Report NIHR HTA Programme project number 11/118

Technology Assessment Report commissioned by the NIHR HTA Programme on behalf of the National Institute for Health and Care Excellence.

Title: *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)*

Produced by: *Warwick Evidence*
Division of Health Sciences
Warwick Medical School, University of Warwick
Coventry
CV4 7AL

Lead Author: *Aileen Clarke, Professor of Public Health and Health Services Research and Director for Warwick Evidence*

Co-authors: *Ruth Pulikottil-Jacob, Research Fellow* [1]
Amy Grove, Research Project Manager [1]
Karoline Freeman, Research Fellow [1]
Hema Mistry, Assistant Professor [1]
Alexander Tsertsvadze, Senior Research Fellow [1]
Martin Connock, Senior Research Fellow [1]
Rachel Court, Information Specialist [1]
Ngianga-Bakwin Kandala, Principal Research Fellow [1]
Matthew Costa, Professor of Trauma and Orthopaedics [1]
Gaurav Suri, Research Associate [1]
David Metcalfe, Academic Clinical Fellow [2]
Michael Crowther, Research Associate in Medical Statistics [3]
Sarah Morrow, Associate Medical Student [4]
Samantha Johnson, Information Specialist [1]
Paul Sutcliffe, Associate Professor [1]

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

Report NIHR HTA Programme project number 11/118

Correspondence to: *Prof Aileen Clarke, Warwick Evidence, Warwick Medical School, University of Warwick, Coventry, CV4 7AL*
Tel: +44 (0) 24761 50063
Email: Aileen.Clarke@warwick.ac.uk

Date completed *9 July 2013*

Source of funding: *This report was commissioned by the NIHR HTA Programme as project number 11/118.*

Declared competing interests of the authors

The authors have no conflicts of interest.

Acknowledgements

The authors would like to thank the National Joint Registry for England and Wales 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, provision of methodological, policy and clinical perspectives on data and results and critique of the report drafts. Clinical advisors include: Professor Ashley Blom-Head of the Orthopaedic Group of the University of Bristol and Professor Alistair Hart-University College London Chair of Orthopaedic Surgery Consultant Orthopaedic Surgeon Director of Research & Development. The biomechanical technical advisors during the work were Professor Richie Gill – Professor of Healthcare Engineering University of Bath, and Dr James Meredith-Senior Teaching Fellow WMG, The University of Warwick. Michael Crowther is funded by a NIHR Doctoral Fellowship (DRF_2012-05-409). We would also like to thank Mrs Hannah Fraser for her input in assembling material and formatting the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Clarke A, Pulikottil-Jacob R, Grove A, Freeman K, Mistry H, Tsertsvadze A, Connock M, Court R Kandala N-B, Costa M, Suri G, Metcalfe D, Crowther M, Morrow S, Johnson S and Sutcliffe P, 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). Warwick Evidence, 2013.

Contributions of authors

Paul Sutcliffe (Associate Professor), Alex Tsertsvadze (Senior Research Fellow) Karoline Freeman (Research Fellow) Amy Grove (Research Project Manager) and Aileen Clarke (Professor of Public Health & Health Services Research) coordinated and conducted the clinical, cost-effectiveness and registries systematic reviews, this included: screening and retrieving papers, assessing against inclusion criteria, appraising the quality of papers and abstracting data from papers for synthesis. Professor Matt Costa (Professor of Trauma and Orthopaedic Surgery) provided clinical comment on the draft and input into the data analysis and formation of the categories. Amy Grove, David Metcalfe (Academic Clinical Fellow), Alex Tsertsvadze and Sarah Morrow (medical student) wrote the background section of the report. Gaurav Suri (Research Associate), Ngianga-Bakwin Kandala (Principal Research Fellow) performed the database analysis. Rachel Court (Information specialist) and Samantha Johnson (Information specialist) developed the search strategy and undertook searches. Martin Connock (Senior Research Fellow), Kandala Ngianga-Bakwin (Principal Research Fellow and Michal Crowther (Research Associate) conducted the survival analysis analysed the data and developed transition probabilities. Ruth Pulikottil Jacob (Research Fellow), Martin Connock, Hema Mistry (Assistant Professor), Paul Sutcliffe and Aileen Clarke analysed the PROMs data and provided advice on the economic model. Aileen Clarke managed design and analysis, and wrote the abstract, summary and discussion. Amy Grove coordinated the project. All authors were involved in writing draft and final versions of the report.

Please refer to the International Committee of Medical Journal Editors (ICMJE) ***Uniform Requirements for Manuscripts Submitted to Biomedical Journals see*** *http://www.icmje.org/*

Report NIHR HTA Programme project number 11/118

Table of contents

1 DEFINITION OF TERMS AND LIST OF ABBREVIATIONS 8

2 EXECUTIVE SUMMARY 17
2.1 Background 17
2.2 Decision problem and objectives 18
2.3 Systematic reviews 18
2.3.1 Systematic review methods 18
2.3.2 Systematic review results 19
2.4 Cost-effectiveness 20
2.4.1 Cost-effectiveness methods 20
2.4.2 Cost-effectiveness results 21
2.5 Strengths and limitations 23
2.6 Conclusions 24
2.6.1 Systematic reviews 24
2.6.2 RS vs. THR 25
2.6.3 THR vs. THR 25
2.7 Recommendations for research 25

3 BACKGROUND 26
3.1 Description of the health problem 26
3.1.1 Aetiology, pathology and prognosis 26
3.1.2 Impact of the health problem 29
3.1.3 Current service provision 30
3.1.4 Relevant national guidance 33
3.2 Description of technology under assessment 36
3.2.1 Summary of THR 36
3.2.2 Summary of hip resurfacing arthroplasty (RS) 51
3.2.3 Failure of hip replacement 52
3.2.4 Revision of hip arthroplasty 54
3.3 Current usage in the NHS 55
3.3.1 General statistics 55
3.3.2 Hip replacement surgery 55
3.3.3 Type of procedure 56
3.4 Background summary 56

4 DEFINITION OF THE DECISION PROBLEM 59
4.1 Decision problem 59
4.2 Overall aims and objectives 59

5 JOINT REGISTRIES 62
5.1 Description of the three largest international registries 62
5.1.1 Australian Orthopeadic Association National Joint Registry 63
5.1.2 The Swedish Hip Arthroplasty Register 63
5.1.3 National Joint Registry for England and Wales 64
5.2 Summary of national registries 65

6 ASSESSMENT OF EVIDENCE 66
6.1 Clinical effectiveness methods 66
6.1.1 Identification of studies 66
6.1.2 Search strategies 66
6.1.3 Inclusion criteria 68
6.1.4 Exclusion criteria 69

6.1.5 Study selection process *69*
6.1.6 Quality assessment strategy *69*
6.1.7 Grading overall quality of clinical effectiveness evidence *70*
6.1.8 Data extraction strategy *71*
6.1.9 Data management *72*
6.1.10 Meta-analysis *72*
6.1.11 Publication bias *73*
6.1.12 Analysis to explore heterogeneity *73*
6.1.13 Data synthesis and interpretation *73*
6.1.14 Industry submissions regarding effectiveness of treatments *74*
6.2 Clinical effectiveness results 74
6.2.2 Comparison of total hip replacement (THR vs. THR) *77*
6.2.3 Grading overall quality of evidence *104*
6.2.4 Summary conclusions comparing THRs *106*
6.2.5 Comparison of total hip replacement and resurfacing arthroplasty (THR vs. RS) *109*
6.2.6 Grading overall quality of evidence *121*
6.3 Summary conclusions comparing THR and RS 123
6.4 Overall summary of clinical effectiveness findings 124
6.5 Cost-effectiveness methods 128
6.5.1 Identification of studies *128*
6.5.2 Inclusion criteria *129*
6.5.3 Assessment of eligibility *130*
6.5.4 Data extraction *130*
6.5.5 Quality assessment *131*
6.6 Cost-effectiveness results 131
6.6.1 Identification of studies *131*
6.6.2 Description of included studies *134*
6.7 Summary of overall cost-effectiveness evidence 148
6.8 Registries methods 149
6.8.1 Identification of studies *149*
6.8.2 Inclusion and exclusion criteria *149*
6.8.3 Assessment of eligibility *150*
6.8.4 Data extraction *151*
6.9 Results of registry review 151
6.9.1 Identification of studies *151*
6.9.2 Review of included studies following stage two exclusion criteria *153*
6.10 Summary of overall registry evidence 167

7 INDIVIDUAL PATIENT DATASET 168
7.1 Introduction to IPD analysis 168
7.2 Method 168
7.3 Selection of patients 168
7.4 Structure of the database 168
7.5 Contents of the database 169
7.6 Results 169
7.6.1 Hip RS arthroplasty *169*
7.6.2 Total hip replacement *171*
7.7 THR category development 171
7.8 Matching 176
7.9 Assessment of utility and quality of the NJR database 176
7.10 Summary of individual patient dataset 177

8 PATIENT REPORTED OUTCOME MEASURES 178
8.1 Quality of life and utilities 178
8.1.1 Background 178
8.1.2 Methods 178
8.1.3 Results 179
8.1.4 Summary of PROMs 181

9 MODELLING REVISION RATES 182
9.1 Introduction 182
9.2 Revision rates 183
9.2.1 Categories of THR 183
9.2.2 Patient populations to be compared 183
9.2.3 Overall revision rates, competing risks (CR) and rationale for analysis 188
9.3 Results 190
9.3.1 Proportional hazards tests 191
9.3.2 Comparison of RS vs. THR 193
9.3.3 Comparison of THR categories 200
9.3.4 Comparison of RS and THR; subgroup analyses according to gender (females) 209
9.3.5 Comparison of RS and THR according to gender (males) 211
9.3.6 Comparison of THR revision rates according to gender and age; men > 65 years old 212
9.3.7 Comparison of THR revision rates according to gender and age; women > 65 years old. 213
9.3.8 Comparison of THR revision rates according to gender and age; men < 65 years old 215
9.3.9 Comparison of THR revision rates according to gender and age; women < 65 years old. 216
9.3.10 Comparison of revision rates with NICE benchmark 217
9.4 Flexible parametric modelling 218
9.5 Discussion, methods of modelling revision rates 219
9.5.1 Summary 222

10 WARWICK ECONOMIC ASSESSMENT 224
10.1 Methods 224
10.1.1 De novo analysis 224
10.1.2 Model structure 224
10.1.3 Base-case analyses 227
10.1.4 Structural model assumptions 227
10.1.5 Estimation of model parameters 229
10.1.6 A summary of the transition probabilities, utilities and cost inputs to the cost-utility model 234
10.1.7 Cost-effectiveness analysis 236
10.1.8 Sensitivity analyses 237
10.2 Results of cost-effectiveness analysis 243
10.2.1 The cost-effectiveness analysis results: RS vs. THR 243
10.2.2 The cost-effectiveness analysis results: comparison of THR categories 247
10.3 Sensitivity analysis results 251
10.3.1 Subgroup analyses: RS vs. THR 251
10.3.2 Subgroup analyses: THR vs. THR (over 65's) 255
10.3.3 Subgroup analyses: THR vs. THR (under 65's) 261
10.3.4 Sensitivity analyses: time to revision (bathtub model adjusted for age and gender) 271
10.3.5 Sensitivity analyses: time to revision (lognormal model) 277
10.3.6 Sensitivity analyses: time to revision (lognormal model adjusted for age and gender) . 280
10.3.7 Sensitivity analyses: costs 283
10.3.8 Sensitivity analyses: utilities 291
10.3.9 Discussion of economic assessment 294

10.3.10 Summary of results *295*
10.3.11 Conclusion of cost-effectiveness analysis *298*
10.4 Comparison of results with TA2, TA44, Manufacturer's submission, & international registries. 299
10.5 Summary and critique of manufacturer's submissions 303
10.5.1 DePuy International Ltd *303*
10.5.2 The base case results *311*
10.5.3 Smith & Nephew, Inc *321*
10.5.4 Stryker *323*
10.5.5 JRI Orthopaedics Limited *324*

11 Discussion 327
11.1 Decision problem and objectives 327
11.2 Methods and summary of findings 327
11.3 Strengths and limitations 331

12 CONCLUSIONS AND IMPLICATIONS FOR PRACTICE 335
12.1 RS vs. THR 335
12.2 THR vs. THR 335
12.3 Recommendations for research 335

13 REFERENCES 337

1 DEFINITION OF TERMS AND LIST OF ABBREVIATIONS

Technical terms and abbreviations are used throughout this report.

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
AAOS	American Academy of Orthopaedic Surgeons
AIMS	Arthritis Impact Measurement Scale
AOANJRR	Australian Orthopaedic Association National Joint Replacement Registry
ASA	American Society of Anaesthesiologists
BHR	Birmingham Hip Resurfacing
BIC	Bayesian Information Criterion
BMI	Body mass index
C	Ceramic
Ce	Cemented (hip fixation)
CE	Conformité Européenne
CEAC	Cost-effectiveness acceptability curve
CePoC	Cemented polyethylene (poly) cup on ceramic head
CePoM	Cemented polyethylene (poly) cup on metal head
CC	Comorbidities
CENTRAL	Cochrane Central Register of Controlled Trials
CG	Clinical guideline
CI	Confidence interval
CeL	Cementless (hip fixation)
CeLCoC	Cementless HA coated metal cup (ceramic liner) on ceramic head
CeLPoM	Cementless HA coated metal cup (polyethylene liner) on metal head
CONSORT	CONSOLIdated Standards of Reporting Trials
CR	Competing risk
CRD	Centre for Reviews and Dissemination
CT	Computed Tomography
CoC	Ceramic-on-ceramic articulation
CoCr	Cobalt Chrome
CoP	Ceramic-on-polyethylene articulation
DARE	Database of Abstracts of Reviews of Effects
DSU	Decision Support Unit
EED	Economic Evaluation Database

EQ-5D	EuroQoL 5-Dimensions
EP	Evidence Profile
FDA	US Food and Drug Administration
GRADE	Grading of Recommendations, Assessment, Development, and Evaluation
HA	Hydroxyapatite
HSCI	Health Service Cost Index
HHS	Harris Hip Score
HSRProj	Health Services Research Projects in Progress
HTA	Health Technology Appraisal
HR	Hazard Ratio
HRG4	Healthcare Resource Group v.4
Hy	Hybrid (fixation)
HyPoM	Cementless HA coated metal cup (polyethylene liner) on metal head with cemented stem
HOOS	Hip Disability and Osteoarthritis Outcome Score
HXLPE	Highly Cross-Linked Polyethylene
ICER	Incremental cost-effectiveness ratio
IPD	Individual Patient Data
ISTCs	Independent sector treatment centres
JRI	Joint Replacement Instrumentation
KM	Kaplan-Meier
LISOH	Lequesne Index of Severity for Osteoarthritis of the Hip
LOS	Length of stay
LYG	Life Year Gain
MA	Meta-analysis
MACTAR	McMaster Toronto Arthritis patient preference questionnaire
MCID	Minimal Clinically Important Difference
MD	Mean Difference
MHRA	UK Medicines and Healthcare Products Regulatory Agency
M	Metal
MO	Months
MoM	Metal-on-metal articulation
MoP	Metal-on-polyethylene articulation
MRI	Magnetic Resonance Imaging
NHS	National Health Service

NCC-CC	National Collaborating Centre for Chronic Conditions
NSRC	National Schedule Reference Costs
NICE	National Institute for Health and Care Excellence
NHP	Nottingham Health Profile questionnaire
NJR	National Joint Registry for England and Wales
NR	Not Reported
NN	Nearest Neighbour
NNT	Number Needed to Treat
NS	Statistically Not Significant
NSAIDs	Non-steroidal anti-inflammatory drugs
OA	Osteoarthritis
ODEP	Orthopaedic Data Evaluation Panel
OHS	Oxford Hip Score
OR	Odds Ratio
PbR	Payment by results
P(E)	Polyethylene
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PPB	Parts Per Billion
PROs	Patient Reported Outcomes
PROMs	Patient Reported Outcome Measures
PSS	Personal and Social Services
QALY	Quality-Adjusted Life Year
QOL	Quality Of Life
RA	Rheumatoid Arthritis
RCT	Randomised Controlled Trial
RR	Rate Ration (Relative Risk)
RS	Resurfacing Arthroplasty
ROB	Risk of Bias
SD	Standard Deviation
SF-36	Short Form 36 Health Survey
SHAR	Swedish Hip Arthroplasty Register
SMR	Standardised mortality ratio
SR	Systematic Review
SROB	Summary Risk of Bias

SSI	Surgical Site Infection
SS	Statistically Significant
TA	Technology Appraisal
TENS	Transcutaneous Electrical Nerve Stimulation
THR	Total hip replacement
UCLA	University of California Los Angeles
UHCW	University Hospitals Coventry and Warwickshire
UHMW	Ultra-high molecular weight
UKCRN	UK Clinical Research Network
Vs.	Versus
WOMAC	Western Ontario McMaster Osteoarthritis Index
WTP	Willingness-to-pay
XLPE	Cross-Linked Polyethylene
YR(s)	Year(s)

LIST OF TABLES

Table 1. THR and RS articulation and fixation type combinations 43
Table 2. Percent spread of fixation type and bearing/articulation surface for primary hip replacements reported in NJR 2011 44
Table 3. Combination of bearing/articulation surface and fixation method by frequency as reported in NJR 2010-11 44
Table 4. Revision procedures by type and year as published NJR data 54
Table 5. Percentage of procedures by organisation type reported in the NJR 2010/11 55
Table 6. Percentage of intervention by fixation method across NHS hospitals and treatment centres reported in the NJR 2010/11 56
Table 7. PICO table 60
Table 8. Joint (hip) replacement registries available worldwide 62
Table 9. Overall study characteristics across 13 RCTs comparing THRs 78
Table 10. Distribution of 13 RCTs according to basis of THR comparison 78
Table 11. Risk of bias for RCTs: review author's judgments about each risk of bias item (THR vs. THR) 80
Table 12. Methodological quality assessment summary for systematic reviews (THR vs. THR) 82
Table 13. Harris Hip score (range: 0-100) - RCTs 85
Table 14. Harris Hip score (range: 0-100) - Systematic reviews 87
Table 15. The Western Ontario and McMaster University Osteoarthritis Index (range: 0-100) - RCTs 88
Table 16. The McMaster-Toronto Arthritis Patient Preference Disability Questionnaire score (range: 0-30) - RCTs 89
Table 17. Merle D'Aubigne and Postel score (range: 0-18) - RCTs 89
Table 18. The University of California, Los Angeles activity scale (range: 1-10) - RCTs 90
Table 19. Oxford Hip score (range: 0-48) - Systematic reviews 90
Table 20. Short Form Health Survey (SF-12; range: 0-100) - RCTs 91
Table 21. Short Form Health Survey (SF-12; range: 0-100) - Systematic reviews 92
Table 22. Revision rate (n/N) - RCTs 92
Table 23. Revision rate (n/N) - Systematic reviews 94
Table 24. Mortality rate (n/N) - RCTs 95
Table 25. Femoral head penetration rate (mm/year) - RCTs 97
Table 26. Implant dislocation rate (n/N) – RCTs 98
Table 27. Implant dislocation rate (n/N) - Systematic reviews 99
Table 28. Osteolysis (n/N) - RCTs 100
Table 29. Osteolysis (n/N) - Systematic reviews 101
Table 30. Aseptic loosening (n/N) - RCTs 102
Table 31. Femoral fracture (n/N) - RCTs 102
Table 32. Infection (n/N) - RCTs 103
Table 33. Deep vein thrombosis (n/N) - RCTs 103
Table 34. Aseptic loosening (n/N) - Systematic reviews 104
Table 35. GRADE evidence profile for gradable outcomes reported in RCTs of THR 105
Table 36. Summary of evidence regarding the differences between the different types of THR for each reported outcome (RCTs) 107
Table 37. Summary of evidence regarding the differences between the compared types of THR for each reported outcome (systematic reviews) 109
Table 38. Overall study characteristic across three RCTs comparing THR vs. RS 110
Table 39. Risk of bias summary for RCTs: review author's judgements about each risk of bias item (THR vs. RS) 111
Table 40. Methodological quality assessment summary for systematic reviews (THR vs. RS) 113
Table 41. Summary results for total hip replacement vs. resurfacing arthroplasty – RCTs 116

Table 42. Summary results for resurfacing arthroplasty vs. total hip replacement – Systematic reviews 120
Table 43. GRADE evidence profile for gradable outcomes reported in RCTs of THR vs. RS 122
Table 44. Summary of evidence regarding the differences between THR and RS for each reported outcome in RCTs 123
Table 45. Summary of evidence regarding the differences between THR and RS for each reported outcome in systematic reviews 124
Table 46 Critical appraisal of the economic evaluation studies using the CHEC-list 139
Table 47 Critical appraisal of the economic models using an adapted Philips checklist 140
Table 48. Characteristics of key cost-effectiveness studies informing the Markov model 144
Table 49. Summary table of registry studies for RS 157
Table 50. Summary table of registry studies for THR 164
Table 51. Total number of male gender by head size for RS 170
Table 52. Total number of female gender by head size for RS 170
Table 53. Characteristics of the five THR categories 174
Table 54. Constitution of THR categories by age and gender 174
Table 55. Prosthesis characteristics for the five categories of THR 175
Table 56. Age and gender of RS and THR recipients 185
Table 57. Age and gender charateristics of patient groups receiving THR protheses 188
Table 58. Age and gender mix of RS and THR populations 196
Table 59 Information criteria scores for models of revision rates (RS and matched THR) 198
Table 60. Bath tub modelled percentage of patients requiring revision 199
Table 61. Information criteria scores for models of revision rates (THR categories) 202
Table 62. Bath tub modelled percentage of patients requiring revision 205
Table 63. Age and gender of recipients of THR categories A to E 205
Table 64. Lognormal modelled percentage of patients requiring revision 208
Table 65. Bath tub modelled percentage of patients requiring revision (females aged 53.5) 210
Table 66. Bath tub modelled percentage of patients requiring revision (males aged 57.1 years) 212
Table 67. Lognormal modelled percentage of patients requiring revision (males > 65 years) 213
Table 68. Lognormal modelled percentage of patients requiring revision (females > 65 years) 214
Table 69. Bath tub modelled percentage of patients requiring revision (males < 65 years) 216
Table 70. Modelled percentage of patients requiring revision (females < 65 years) 217
Table 71. Estimated percentage of patients requiring revision at 10 years 218
Table 72. Key features of the analysis 226
Table 73. Summary of assumptions 228
Table 74. RS prosthesis cost as reported by the NHS supply chain 230
Table 75. Cost of THR prosthesis 231
Table 76. Total cost of surgery 231
Table 77. Cost of revision 232
Table 78. Cost of successful revision procedure (THR/RS) 232
Table 79. Prosthesis cost 233
Table 80. Cost of bone cement pack 234
Table 81. Bathtub parameters for comparison RS vs. THR and THR vs. THR 235
Table 82. Summary of transition probabilities, utilities and cost inputs for base-case analysis 235
Table 83. Sub-group analysis – Time to revision for RS vs. THR and THR vs. THR 238
Table 84. Summary of utilities inputs for sub-group analysis 238
Table 85. Sensitivity analysis – Time to revision for RS vs. THR 239
Table 86. Sensitivity analysis – Time to revision for THR vs. THR 240
Table 87. Highest and lowest list price for both RS and THR (weighted average of all categories) prosthesis 241
Table 88. Prosthesis cost for sensitivity analysis – highest and lowest unit costs and using a 20% reduction in prosthesis list price across all categories 241

Table 89. Summary of utilities inputs for sensitivity analysis 242
Table 90. Base-case deterministic and probabilistic results for all patients using bathtub model 243
Table 91. Base-case deterministic and probabilistic results for all THR patients using bathtub model ... 247
Table 92. Deterministic results for 40, 50 and 60 year old males and female patients 251
Table 93. Probabilistic results for 40, 50 and 60 year old male and female patients 252
Table 94. Deterministic results for males and females over 65 years of age for a 10-year time horizon . 255
Table 95. Probabilistic results for males and females over 65 years of age for a 10-year time horizon... 256
Table 96. Deterministic results for males and females over 65 years of age for a lifetime horizon 257
Table 97. Probabilistic results for males and females over 65 years of age for a lifetime horizon 258
Table 98. Deterministic results for males under 65 years of age for 10-year time horizon 261
Table 99. Probabilistic results for males under 65 years of age for 10-year time horizon..... 262
Table 100. Deterministic results for males under 65 years of age for a lifetime horizon 263
Table 101. Probabilistic results for males under 65 years of age for a lifetime horizon..... 264
Table 102. Deterministic results for females under 65 years of age for a 10-year time horizon 267
Table 103. Probabilistic results for females under 65 years of age for a 10-year time horizon..... 268
Table 104. Deterministic results for females under 65 years of age for a lifetime horizon 269
Table 105. Probabilistic results for females under 65 years of age for a lifetime horizon..... 270
Table 106. Deterministic and probabilistic results for all patients using bathtub model adjusted for age and gender 271
Table 107. Deterministic and probabilistic results for all THR patients using bathtub model adjusted for age and gender 274
Table 108. Deterministic and probabilistic results for all THR patients using lognormal model..... 277
Table 109. Sensitivity analysis deterministic and probabilistic results for all THR patients - age and gender adjusted using a lognormal model 280
Table 110. Deterministic and probabilistic results for lowest and highest costs for THR vs. RS patients 283
Table 111. Deterministic and probabilistic results using the highest prices for all THR patients using a bathtub model 286
Table 112. Deterministic and probabilistic results using the lowest prices for all THR patients using a bathtub model 287
Table 113. Deterministic and probabilistic results assuming a price de-escalator of 20% for all THR patients using a bathtub model 288
Table 114. Deterministic and probabilistic using utility values from Rolfson 291
Table 115. Patient populations considered in the manufacturer's model based on NJR data..... 307
Table 116. Utility values used in DePuy model..... 307
Table 117. Values for resource costs used in DePuy model 308
Table 118. Base case results of DePuy submission considering THR vs hip resurfacing in patients suitable for both procedures..... 311
Table 119. Base case results for THR categories and RS of DePuy submission for patients suitable for THR and hip resurfacing 311
Table 120. Base case results for THR categories of DePuy submission for patients not suitable for hip resurfacing arthroplasty 312
Table 121. Results reported in the DePuy submission for the scenario analyses of patients suitable for hip resurfacing arthroplasty 313
Table 122. Results reported in the DePuy submission for the scenario analyses of patients not suitable for hip resurfacing arthroplasty 315
Table 123. Results of the probabilistic sensitivity analysis by DePuy for patients suitable for THR and hip resurfacing arthroplasty 316
Table 124. Results of the probabilistic sensitivity analysis by DePuy for patients not suitable for hip resurfacing arthroplasty 316
Table 125. Mean length-of-stay for patients receiving primary THR or hip resurfacing 319
Table 126. Age and gender of patients receiving primary hip replacements in 2011 320

LIST OF FIGURES

Figure 1. Example pathway for patient with arthritis in primary care 39
Figure 2. Example hip replacement care pathway in secondary care 40
Figure 3. Overview of four different fixation options for the femoral stem and acetabular cup in THR ... 42
Figure 4. Cemented metal stem, metal femoral head and polyethylene acetabular cup 45
Figure 5. Cemented metal stem, ceramic femoral head and polyethylene acetabular cup..... 46
Figure 6. Cemented metal stem, ceramic femoral head and ceramic acetabular cup..... 47
Figure 7. Cemented metal stem, metal femoral head and metal acetabular cup 48
Figure 8. Diagrammatic representation of a hip resurfacing arthroplasty (RS) 51
Figure 9. PRISMA 2009 flow diagram of clinical effectiveness 76
Figure 10. Risk of bias graph for RCTs: review author's judgments about each risk of bias item (THR vs. THR) 81
Figure 11. Harris Hip Score 86
Figure 12. Mortality 96
Figure 13. Implant dislocation 99
Figure 14. Risk of bias graph for RCTs: review author's judgements about each risk of bias item (THR vs. RS) 112
Figure 15. Risk of infection 118
Figure 16. Risk of deep vein thrombosis 118
Figure 17. Risk of implant dislocation 118
Figure 18. PRISMA flow diagram cost-effectiveness studies 133
Figure 19 PRISMA flow diagram for registries studies..... 153
Figure 20. Endpoint for all RS included in the analysis 170
Figure 21. End point of all THR included in the analysis..... 171
Figure 22. The frequency of each of the five categories of the THR dataset..... 173
Figure 23. Age distributions of NJR patients receiving THR or RS, and by gender for RS..... 184
Figure 24. Kernel density diagram of the two distributions..... 185
Figure 25. Differing modelled hazard in extrapolation beyond observation (HyPoM [Category D] THR females <65 years old) 189
Figure 26. Kaplan Meier analysis for death (left) and revision (right) for THR CePoM (Category A) female patients > 85 years old..... 190
Figure 27. Log KM estimated cumulative hazard versus log time. 191
Figure 28. Log K-M estimated cumulative hazard versus log time for different THR categories 192
Figure 29. Log K-M estimated cumulative hazard versus log time for different THR categories 192
Figure 30. Time to revision; all RS patients and all THR patients 193
Figure 31. Time to revision; all RS and all THR patients according to gender 193
Figure 32. Time to revision all RS and all THR patients (Categories A to E)..... 194
Figure 33. Time to revision all RS and all THR patients (Categories A to E) by gender..... 194
Figure 34. Time to revision all male RS and all male THR patients (Category A to E)..... 195
Figure 35. Time to revision all female RS and all female THR patients (Category A to E) 195
Figure 36. Age distribution and Time to revision for RS and THR matched populations..... 196
Figure 37. Revision rates for matched and whole THR populations 197
Figure 38. Bath tub fits and extrapolations for matched THR and RS populations..... 198
Figure 39. K-M versus modelled cumulative hazard 199
Figure 40. Revision estimated for all THR patients and those receiving category A to E THRs 200
Figure 41. Observed time to revision; upper panel THR categories compared; lower panel THR categories with 95% CI; cement-less upper row, cemented lower row..... 201
Figure 42. K-M versus modelled cumulative hazard 203
Figure 43. Bath tub parametric fits to observed time to revision for THR categories A to E..... 204
Figure 44. Extrapolation of bath tub models of revision for THR categories A to E 204

Figure 45. THR revision rates observed for males and females: all THR categories 206
Figure 46. Kernel density plots of age at primary for category A to E THR prostheses 206
Figure 47. Lognormal parametric fits to observed revision for THR categories A to E 207
Figure 48. Lognormal modelled revision; uncontrolled (left), controlled for population of mean age 71.6 years and 36.5% male 208
Figure 49. Kernel density plot for age distribution in matched RS and THR female groups 209
Figure 50. Observed revision (95% CI) & bath tub models for RS & THR 210
Figure 51. Kernel density plot for age distribution in matched RS and THR male groups 211
Figure 52. Observed revision (95% CI) and bath tub models for RS & THR males 212
Figure 53. Observed revision (95% CI) for males > 65 years old & lognormal models for THR categories 213
Figure 54. Observed revision (95% CI) for females > 65 years old & lognormal models for THR categories 214
Figure 55. Observed revision (95% CI) for males < 65 years old & bath tub models for THR categories 215
Figure 56. Observed revision (95% CI) for females < 65 years old & bath tub models for THR categories 217
Figure 57. Markov model 225
Figure 58. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. RS 246
Figure 59. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. THR using a bathtub model 250
Figure 60. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. RS for females by age group (lifetime horizon) 253
Figure 61. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. RS for males by age group (lifetime horizon) 254
Figure 62. Cost-effectiveness acceptability curves for THR vs. THR (over 65s) 260
Figure 63. Cost-effectiveness acceptability curves for THR vs. THR for females and males by age group (lifetime horizon) 266
Figure 64. Cost-effectiveness acceptability curves for RS vs. THR age and gender adjusted using a bathtub model 273
Figure 65. Cost-effectiveness acceptability curves for RS vs. THR age and gender adjusted using a bathtub model 276
Figure 66. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. THR using a lognormal model 279
Figure 67. Cost-effectiveness acceptability curves for THR vs. THR age and gender adjusted using a lognormal model 282
Figure 68. Cost-effectiveness acceptability curves for RS vs. THR using lowest and highest costs 285
Figure 69. Cost-effectiveness acceptability curves for all THR patients using lowest and highest costs and assuming a 20% price de-escalator 290
Figure 70. Cost-effectiveness acceptability curves for all THR using utility values 292
Figure 71. Tornado diagram illustrating sensitivity analysis for net monetary benefit: CePoP versus CePoM 293
Figure 72. Swedish Registry data for time to revision to 19 years of follow up 301
Figure 73. Time to revision results for Schulte, Madey, and Callaghan, respectively 302

2 EXECUTIVE SUMMARY

2.1 Background

Arthritis is a general term describing pain and inflammation within a joint. Osteoarthrisits (OA) is a leading cause of pain and disability both in the UK and worldwide. It is a chronic disorder of articular cartilage degeneration. 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. The prevalence and incidence of OA increase with age and both are higher in women than men after 50 years of age. The economic impact of arthritis is vast, both due to direct costs to the healthcare system, community and social services and due to indirect costs due to restricted activity, lost productivity and early mortality. OA in the hip manifests itself as pain on activity and at later stages, at rest. Patients who do not respond to non-surgical measures e.g. analgesia etc. are referred for elective surgical interventions, most commonly total hip replacement (THR) or hip resurfacing arthroplasty (RS).

THR involves replacement of a damaged hip joint with an artificial hip prosthesis consisting of a cup (with or without liner), a femoral stem, and a femoral head. There are different types of THR including different types of articulation surface (metal, ceramic, polyethylene, ceramicised metal); methods of implant component fixation (cemented, cementless, hybrid, reverse hybrid); and implant component size (e.g., femoral head size). Approximately 80,000 hip replacement operations are done a year in the UK of which approximately 95% are for OA. Rates for primary and revision THR have been increasing with a 16% increase recorded between 2005 and 2010. The greatest proportion of procedures (65%) is in patients aged 65 years and older.

RS involves replacement of the joint surface of the femoral with a metal surface covering. The new resurfacing component articulates with a hollow metal cup located in the acetabulum. RS is thought to allow for more bone preservation and lower risk of dislocation compared with standard THR, and to be more suitable for younger and more active patients. Revision surgery using THR undertaken when RS or THR implants fail due to infection, or osteolysis and loosening.

Previous National Institute for Health and Care Excellence (NICE) guidance on this issue including TA2 and TA44 indicated that the benchmark for selection of prostheses for THR should be a revision rate of 10% or less at 10 years. Available evidence supported the use of a range of cemented prostheses for primary THR. In June 2002, NICE Technology Assessment (TA) 44 guidelines recommended, metal on metal (MoM) hip RS arthroplasty as one option for people with advanced hip disease who would

otherwise receive, and are likely to outlive, a conventional primary hip replacement. However, in June 2012 advice about follow-up of patients receiving a MoM articulation changed as a result of research on complications. The Medicines and Healthcare products Regulatory Agency (MHRA) issued a medical device alert and recommended that clinicians should perform appropriate follow-up.

2.2 Decision problem and objectives

The main aim was to undertake a clinical 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 objectives were the following:

To compare the clinical and cost-effectiveness of:

A. Different types of primary THR compared with RS for people in whom both procedures are suitable

B. Different types of primary THR compared with each other for people who are not suitable for hip RS

2.3 Systematic reviews

2.3.1 Systematic review methods

Searches were undertaken of clinical effectiveness, registry and cost-effectiveness studies in December 2012, and limited for clinical effectiveness studies to studies published from 2008 and onward and to a sample size of 100 participants or more. Electronic searches were conducted in: MEDLINE, MEDLINE In-Process, Embase, Science Citation Index, Cochrane Library (Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials), Current Controlled Trials, ClinicalTrials.gov, Database of Abstracts of Reviews of Effectiveness (DARE), and Health Technology Assessment (HTA) database. Reference lists, and websites of hip implant manufacturers, and major orthopaedic professional organizations were screened for relevant publications. Full text English-language reports of Randomised Controlled Trials (RCTs), systematic reviews, and meta-analyses (MA) were included.

Two independent reviewers screened all records, and extracted data and disagreements were resolved through consensus or with a third reviewer. Two reviewers independently assessed risk of bias of RCTs and methodological quality of systematic reviews using the Cochrane Collaboration's Risk of Bias (ROB) tool and the AMSTAR tool. Estimates of post-treatment mean difference (MD) for continuous outcomes

and risk ratios (RR) for binary outcomes (except for rare events) 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 (OR) using the Peto fixed-effects model. Statistical heterogeneity was determined through Cochran's Q and the I^2 statistics according to the pre-determined levels of statistical significance (Chi-square $p < 0.10$ and/or $I^2 > 50\%$). Overall quality of evidence was assessed using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) Working Group approach.

2.3.2 Systematic review results

2.3.2.1 Clinical effectiveness: RS versus THR

A total of 2,469 records were screened of which 37 were included, representing 16 RCTs and 8 systematic reviews. Thirteen RCTs from the USA (n=4), the UK (n=1), Australia (n=1), Norway (n=2), South Korea (n=2), and Canada (n=3) compared different types of THR based on composition, design, bearing surface, fixation method, and size of implant components. Mean age ranged from 45 to 72 years with maximum follow up of 20 years. Mean post-THR Harris Hip score measured at different follow-ups (6 months to 10 years) did not differ between THR groups including between cross-linked polyethylene vs. traditional polyethylene cup liner (pooled mean difference=2.29, 95% Confidence Interval (CI): -0.88, 5.45). Similarly, there were no differences in WOMAC and SF-12 scores between THR groups of cross-linked polyethylene vs. traditional polyethylene cup liners.

There was a reduced risk of implant dislocation with the use of cemented cup vs. cementless cup; high grade; (pooled OR=0.34, 95% CI: 0.13, 0.89) or larger femoral head size (36 mm versus (vs.) 28 mm). Patients who received THR with cross-linked compared to conventional polyethylene cup liners experienced reduced femoral head penetration rate and risk of revision but this latter finding was very low evidence grade. Recipients of ceramic-on-ceramic articulations (vs. metal-on-polyethylene) experienced a reduced risk of osteolysis.

Five systematic reviews reported evidence on types of THR (cemented vs. cementless cup fixation and implant articulation materials), but these reviews were inconclusive due to unreported pooled results, inappropriate pooling methods or inconsistent summary findings.

2.3.2.2 Cost-effectiveness reviews

1,664 records were screened. Sixty-six studies were included in the narrative review and four of the 11 core studies identified provided relevant date for the model in terms of costs and utilities.

2.3.2.3 Registry review

541 records were screened. Thirty relevant studies representing registries in Scandinavia, England and Wales and Australasia were included. No studies reported better implant survival for RS than for all THR. One study of males reported that RS had a similar revision rate to an uncemented THR, but that both had a higher revision rate than a cemented THR. Three studies from the Swedish joint registry and the National Joint Registry (NJR) informed the survival analysis.

2.4 Cost-effectiveness

2.4.1 Cost-effectiveness methods

For both research questions we drew on systematic review results and we used the NJR to identify populations undergoing the various types of interventions. Having identified a group undergoing RS, we subdivided the group undergoing THR. Using a series of cross tabulations, we identified the top four most commonly used mutually exclusive categories of THR (>25,000) and our clinical advisors recommended inclusion of a further mutually exclusive 5th category. We investigated observed time to revision for all RS, for all THR, for all of our identified categories of THR combined and for each of our categories separately using NJR data. We investigated a number of methods for extrapolating beyond observed data and tested goodness of fit.

We built a Markov multi-state model to investigate both RS and THR. Health states included successful primary surgery, revision surgery, successful revision surgery and death. Cycle length was one year. We adopted a 10-year and a lifetime horizon. The analysis was conducted from the perspective of the NHS and personal and social services (PSS). All costs are in pounds sterling (£) at 2011/2012 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. We calculated cost-effectiveness acceptability curve (CEACs) and undertook sensitivity analyses.

2.4.1.1 RS vs. THR

We propensity-matched RS patients drawing age-gender matched pairs from the dataset of all identified categories of THR combined, in order to identify patients who underwent THR but who were also eligible for RS. We used NHS Supply Chain costs for both RS and THR for follow up and literature sources for revision costs. We drew age-and gender-adjusted utility values from the patient reported outcome measures (PROMs) dataset for both THR and RS.

For the comparison of RS versus THR we undertook sensitivity analyses stratified by gender 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 and in separate age groups at different levels of willingness to pay.

2.4.1.2 THR vs. THR

We compared the five categories of THR with each other, investigating patients eligible for THR (all patients) and those aged >65 years who are less eligible for RS in sensitivity analyses. For the base case we used NHS Supply Chain costs (cup, liner, head, stem and coating) including both cemented and cementless options where appropriate. We used highest and lowest costs supplied in sensitivity analyses. We used age and gender adjusted utility values from the PROMs dataset for before and after hip replacement and for revision.

We undertook sensitivity analyses and analysis of cost drivers including investigating age and gender categories, stratifying by age (less than and more than 65 years), different methods of extrapolation of revision rates (using a lognormal model) and by varying prosthesis costs and discount rates. We constructed CEACs comparing different types of THR overall and in separate age groups at different levels of willingness to pay.

2.4.2 Cost-effectiveness results

2.4.2.1 NJR and PROMs data

Using the NJR we found a total of 31,222 people who had undergone a hip RS procedure and 386,556 undergoing a THR. Our identified categories of THR covered 62% of the THR population. We found that populations undergoing RS and THR overlapped substantially (for RS 89.7% were less than 65 years old and for all THR categories 22.6% were less than 65 years old). We found that for extrapolation, bathtub models (indicating increasing likelihood of revision over time) gave the best fit to observed data. PROMs data showed that utility differences were dramatic – i.e. from pre intervention at 0.35 to post intervention at 0.78 and from pre-revision at 0.53 to a similar level for post revision at 0.78.

2.4.2.2 RS vs. THR

Using the NJR data, we found that 97% of those undergoing THR had remained without a revision by 9 years, whereas only 89% of those undergoing RS had not had a revision by the same stage.
We found that revision rates for all RS, compared to THR (all THR, all of our identified categories of THR combined, each of our THR categories separately) were always higher. For that reason we

undertook our modelled comparisons of cost-effectiveness using all our THR categories combined vs. RS. The costs of the prostheses were; RS £2,672 and THR £2,571.

For all analyses, mean costs for RS were higher than THR; and 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 total 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 analysed separately in sensitivity analyses for males and females by age group (40, 50 and 60 years). For all age and gender groups RS remained clearly dominated by THR. CEACs showed that for all patients, THR was almost 100% cost-effective at any willingness to pay level.

2.4.2.3 THR vs. THR

Given the lack of high quality RCT evidence for different types of THR we used the NJR as our major source of information. We identified five categories of commonly used types of THR:

- Category A: CePoM (Cemented-cemented with a polyethylene-metal articulation) (125,285 patients)
- Category B: CeLPoM (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: HyPoM (Hybrid (cementless-cemented) with a polyethylene-metal articulation) (28,471 patients) and
- Category E: CePoC (Cemented-cemented with a polyethylene-ceramic articulation) (12,075 patients)

There were age and gender differences in the populations with 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). Although for all interventions, revision rates at nine years were substantially less than the 10% benchmark. Costs of the different prostheses were as follows: Category A – CePoM £1,557.38; B – CeLPoM £3,015.60; C – CeLCoC £3,868.80; D – HyPoM £2,649.78 and E – CePoC £1,995.98.

For the base-case analysis, for all age and gender groups combined and using a bathtub model (indicating increasing likelihood of need for revision with time), and a lifetime horizon, mean costs for Category E (CePoC) were slightly lower and mean QALYs for Category E were slightly higher, than for all other THR categories in both deterministic and probabilistic analyses. Hence, Category E dominated the other four categories. For example in the deterministic analysis, compared to Category E, Category A (CePoM) cost £278 more (£14,801 compared to £14,523) and generated 0.0022 fewer QALYs (14.7887 as compared to 14.7909) and 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 to Category A (2.8%) at a willingness to pay of £20,000 per QALY. For patients aged over 65 years, at a willingness to pay of £20,000 per QALY, Category A was more likely to be cost-effective in all groups (Category A: 100% probability of being cost-effective; Categories B,C,D,E: <1% probability of being cost-effective).

Sensitivity analyses using a lognormal model (indicating a decreasing risk of revision over time) for extrapolation beyond the observed data for revision rates, gave category A as cheaper at a lifetime horizon for all age-gender groups combined. Although category E was more effective than the other four categories, Category A was 100% cost-effective at a willingness to pay threshold of £20,000 per QALY. Further sensitivity analysis using an age and gender adjusted log normal model demonstrated the same finding at a lifetime horizon and a willingness to pay of £20,000 per QALY, Category A was 100% cost effective.

Using a one-way sensitivity analysis and varying the main inputs (e.g. costs by 30%) in the base case analysis for all age- gender groups, and comparing Category A with Category E, demonstrated that the main drivers of difference were costs of components, discount rates and modelled revision rates.

2.5 Strengths and limitations

Our literature reviews were rigorous. We reanalysed comprehensive national audit data to calculate outcomes, and used Patient Reported Outcomes Measures (PROMs) as a source for utility data coupled with costs sourced from the literature, from NHS reference costs and from different manufacturers.

However, we did not find any relevant longer term randomised controlled trials covering the comparison between RS and THR or between different types of THR which would allow us to model differences in revision rates for RS or THR relevant to a lifetime horizon. The non-randomised clinical audit data from the NJR may be subject to selection bias. We worked to reduce confounding by propensity-matching RS with THR patients using NJR data and by undertaking extensive analyses by age and gender for the

comparisons of different types of THR. We based our economic model on previous research and a strength is that we had an independent critique and assessment of the model and altered the structure in relation to these external comments.

In comparing RS with THR, our clinical advisors suggested to us that selection of patients for RS may be made based on activity levels (levels of physical fitness, athleticism, weight lifting, manual labour), however the only characteristics which were reliably collected at patient level in the NJR were age and gender. This means that we were unable to identify other characteristics or sub populations where RS might be more beneficial. However age and gender 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 gender group we examined – including in the youngest category of men.

For the comparisons of different types of THR, 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 forces affecting cost-effectiveness of the different categories. We had pre-selected Category E (CePoC) prior to assessing any revision rates on the recommendation of our clinical advisors. We undertook extensive modelling of revision rates in order to find the best methods for extrapolation beyond observed data for all THR categories. We found that Category E had lower revision rates overall and generally across age-gender groups and that this pertained in comparison to other categories across different methods for extrapolation, suggesting that the relative cost-effectiveness of Category E (CePoC) is a robust finding.

2.6 Conclusions

2.6.1 Systematic reviews

THR 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. Overall, revision rates are low. However, although we appraised and summarized a very large amount of evidence much of it was inconclusive due to poor reporting, missing data, inconsistent results and uncertainty in treatment effect estimates. Evidence on the relative benefits of RS vs. THR or of different types of THR was largely lacking. Certain types of THR appeared to confer some benefit included 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.

2.6.2 RS vs. THR

Compared to THR, revision rates for RS were higher, mean costs for RS were higher and mean QALYs gained were lower; RS was dominated by THR.

Very similar results were obtained for deterministic and probabilistic results and for all age and gender groups and THR was almost 100% cost-effective at any willingness to pay level.

2.6.3 THR vs. THR

Revision rates for all types of THR were low. Costs of prostheses varied depending partly on modularity (e.g. presence or absence of a liner etc.) There were small but clear differences between categories in both costs and effectiveness as measured by QALYs and when age and gender groups were factored in. Category A (cemented components with a polyethylene-on-metal articulation) was more cost effective for older age groups where revision rates are lower. However across all age gender groups combined, for the base-case analysis, 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 for all other THR categories in both deterministic and probabilistic analyses; Category E dominated the other four categories.

2.7 Recommendations for research

1. Randomised controlled trials with adequate length of 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 resources use data and implemented routinely would be extremely useful for future cost-effectiveness assessments.

3. We would welcome work to validate our new findings on the relative cost-effectiveness of different combinations of prosthesis components for THR.

3 BACKGROUND

3.1 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]

3.1.1 Aetiology, pathology and prognosis

3.1.1.1 Osteoarthritis of the hip

The hip is a weight-bearing ball and socket joint which 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]

OA 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 where no other specific cause has been identified.[3]

3.1.1.2 Rheumatoid arthritis of the hip

Rheumatoid arthritis (RA) is an autoimmune disease which commonly affects synovial lining of peripheral joints, including those of the hand, foot, and hip. RA is a multi-system 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]

RA 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 developed new criteria for identifying patients with early RA which places more emphasis characteristics associated with a high risk of later progression to severe and erosive disease.[5]

3.1.1.3 Epidemiology of OA and RA

OA 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 over 45 years.[8] A further 8.5 million people are estimated to be affected by joint pain which can be attributed to OA.[3]

Current projections estimate that 10% of the world's population aged 60 years or older 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 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 amongst women and men in Europe are about 53.3 and 38.1 per 100,000.[13] The prevalence of hip OA among Caucasians is demonstrably higher (range 3%-6%) than in Asians, Blacks, and East Indian populations (1% or under).[14] In light of a longer life expectancy an aging population, and increasing rates of obesity observed in the developed countries, it is expected that both the incidence and 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 while 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 is classically between ages 40 and 60. The incidence of RA is approximately

two to three times greater in women than men,[4] and approximately 10-40% of cases manifest within the hip.[19]

3.1.1.4 Risk factors for OA

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

a) Biomechanical factors (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 of OA to affect weight-bearing joints, e.g., hip and knee.[2] Mal-alignment, 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., sports such as 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]

3.1.1.5 Symptoms and diagnosis

Symptoms of hip OA include pain, stiffness, and function, i.e., limited daily activities such as walking, climbing the stairs, 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 due to other causes, e.g., 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 computed tomography (CT) 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]

3.1.1.6 Natural history of OA

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

3.1.2 Impact of the health problem

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

OA has a significant impact on an individual patient with pain, stiffness, limited mobility, and reduced function. A UK based survey assessed the impact of OA on daily living for 1,762 people.[26] The majority of the sample consisted of people aged 50 years or older, of whom 75% were female. Eighty one percent of respondents were found to have experienced constant pain and/or were limited in their ability to perform everyday tasks. Many respondents had visited the 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.

3.1.2.2 Significance for the NHS

The economic impact of arthritis consists of direct costs to healthcare services and indirect costs due to 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 £640.00 per person per year.[19] A recent report has suggested that if one-tenth of the 15.2 people per 1,000 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 £48 million per year in 2002.[19] The costs of both surgical and non-surgical interventions are reviewed in detail below.

Due to 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 healthcare spending, the implications of increasing demand for the treatment of arthritis of the hip have led to intense discussion about the cost-effectiveness of new technologies and treatment options.

3.1.2.3 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 tool 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 Hips 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 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 which 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 d'aubigne and Postel hip score[32] and the Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH).[33-35]

3.1.3 Current service provision

3.1.3.1 Management of disease

Treatment and management of arthritis in the UK can be categorised as non-surgical and surgical as detailed below. Patients at 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:

- Self-management and patient education
- Non-pharmacological (acupuncture, exercise, physical therapy, manual therapy, weight reduction)
- Pharmacological (simple analgesics, non-steroidal anti-inflammatory drugs, topical treatments, intra-articular steroid injections)

Surgical:

- Surgical (e.g., THR or RS, arthrodesis, arthroscopy, osteotomy)

3.1.3.2 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 remains the most common group of conditions for which people receive Disability Living Allowance in England and more than for heart disease, stroke, chest disease, and cancer combined.[36] A reported £43 million is spent annually on community services and £215 million on social services for OA.[36] In 2002 an estimated 36 million workdays were lost due to OA, resulting in £3.2 billion of lost productivity.[36] Data for the numbers of people who have their symptoms managed by non-surgical interventions (such as pain, exercise, physical therapy and manual therapy) within England and Wales are difficult to ascertain.

Chen et al. (2012)[8] estimated the cost of topical and oral non-steroidal anti-inflammatory drugs (NSAIDs) using prescribing data from 2005/06. 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 £8.5 million and £25 million respectively.[8] Adjusting for inflation they found this would equate to £19.2 million and £25.65 million 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 gained.[37]

An earlier Health Technology Assessment (HTA) (reference number 01/21/01) reported that the annual cost to the NHS of elective hip replacement surgery for treatment of OA was £140 million and that each trust spent, on average, £257,000 on the purchase of hip prostheses in 1998/99.[19] The previous HTA was conducted in 2002.[19] It reported the cost to the NHS and social services of non-surgical treatment for an individual to be approximately £640.00 per person per year. During the year 2000, £405 million was spent on 44,000 hip and 35,000 knee replacements.[36] Since then the costs have increased substantially, as the estimated cost of THR surgery alone in the NHS in 2011 was reported to be £426 million.[36]

The cost of one surgical treatment in 2002 was £3,891, averaged across all NHS trusts in 1999/2000, with the cost for 50% of trusts falling within the range £3,404–£4,434.23.[19] According to the 8th annual report from the National Joint Registry[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 2005/06 was £6,000 for a primary THR whereas, in 2010, the national tariff was set at £5,552 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 surgeries has increased in recent years, with 3,012 revision procedures carried out in 2003/4, rising to 6,581 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 for primary surgery. Briggs et al. (2004)[38] reported a mean cost for a standard hip revision procedure in 2000/1 as £5,294 (£6,385; 2008 prices) compared to £3,889 (£4,690; 2008 prices) for a primary procedure. The 2002 HTA reported in 1989/90 that one in seven of all procedures (5,000 out of a total of 35,000) were revisions of hip replacements.[19] In 1999/2000 a crude estimate of 6,700 revisions was reported.

Randomised controlled trials 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 seven year revision rates were lower for cemented (3.0%) than for hybrid (3.8%) or cementless prostheses (4.6%). [36] Edlin et al. (2012)[40] reported that a total of 97% of UK hip replacements are still working (unrevised) at five years.

3.1.3.3 Variation in services and uncertainty about best practice

Outcomes for hip replacement surgery vary by geographical population, 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 £53 million.[43] By 2008 this had risen to 124 brands of acetabular cups and 137 brands of femoral stems at a cost of £67 million.[36] This represents a substantial increase in the variety of available prostheses in recent years. Brands 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 1,000 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 under 55 is 33%.[40]

Spend also varies significantly between regions in the UK, from the lowest reported (£560.00) in Tower Hamlets to the highest in Devon (£8,140).[42] When examining data by local authority, the difference in the rate of provision of hip replacements per 1,000 people in need was almost 14-fold.[42] National EuroQoL 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]

3.1.4 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 osteoarthritis.[3] The National Institute for Health and Clinical Excellence (NICE) developed clinical guidance on the selection of prostheses for primary THR[46] and metal on metal hip RS.[25]

3.1.4.1 Summary of NICE TA2 April 2000-Guidance on the selection of prostheses for primary THR

Technology Assessment 2 (TA) stated that the 'benchmark' for selection of prostheses for THR should be a revision rate of 10% or less 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 post-operative pain.

There were currently no cost-effectiveness data, based on revision rate of 10 years or more 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 prosthesis. 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.

3.1.4.2 Summary of NICE TA44 June 2002 - Guidance on the use of metal on metal hip resurfacing arthroplasty

In June 2002, NICE TA44 guidelines, metal-on-metal (MoM) hip RS arthroplasty 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 less than 65 years of age 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 likely outcome of revision surgery compared with conventional THRs.

However, in June 2012 advice about follow-up of patients receiving a MoM 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 MoM articulations. These reactions could also adversely affect the results of later revision surgery. However, they also stated that their evidence pointed to the fact that early revision of such, poorly performing MoM replacements should give a better revision outcome. Therefore, they advised that clinicians should perform the appropriate follow-up advice, depending on which of the following groups their 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 >7 parts per billion (PPB) indicates potential for soft tissue reaction). Then revision should be considered if imaging was abnormal and/or blood metal ion levels were rising.

3.1.4.3 Summary of MHRA alert advice

MoM 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 implant

MoM THRs with a head diameter <36mm

- 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

MoM THRs with head diameter ≥36mm

- Annual follow-up for life of implant whether symptomatic or not

DePuy ASR hip replacements (all types)

- Annual follow-up for life of implant whether symptomatic or not

3.1.4.4 NICE Guidance on Osteoarthritis February 2008 - The care and management of osteoarthritis in adults

The most recent NICE guidance on osteoarthritis stresses the importance of a holistic assessment of the patient, including their 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 programs 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 non-steroidal anti-inflammatory drugs (NSAIDs) or topical capsaicin (topical treatments less useful for hips though). If these are found 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 already been offered the 'core' treatments and is still experiencing joint symptoms that have a substantial impact on quality of life.

3.1.4.5 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 against 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 (NJR report 2011). ODEP does

have to rely upon the honesty of the submitting companies and therefore make no warranty that the data in their database is accurate, complete or current.

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

- Level A - strong evidence that meets NICE guidance
- Level B - reasonable evidence
- Level C - weak evidence
- Unacceptable evidence

For products that fail to meet NICE's 10 year benchmark, ODEP looks at evidence at three, five and seven years. Again, these are split into whether there exists acceptable, weak or unacceptable evidence for the product meeting NICE guidance.

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 cup and 47% of available brands of femoral stem 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.

3.2 Description of technology under assessment

3.2.1 Summary of THR

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 which 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 for a patient undergoing a hip replacement in 2010 was 67.2 years. There was also a 3% increase in the per cent of females (59%) in 2010/11 compared to 2009. On average, female patients were older than male patients at the time of their THR replacement (68.8 years and 66.3 years, respectively).[36]

Modern THR began in the 1970s with widespread use of the Charnley prosthesis. More than 80,000 procedures are performed every year in England and Wales, with excellent clinical outcomes showing

greater than 95% implant survivorship at 10-year follow-up, and greater than 80% implant survivorship at 25-year 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 in patients aged 65 years and older. However, the proportion of patients undergoing THR who are younger than 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 Section 3.2.1.1.

At 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 non-surgical treatments 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:

- Osteoarthritis (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 due to the risk of poor prosthesis survivorship over the 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.

3.2.1.1 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 recommended for surgery in secondary care.

Figure 2 presents the two surgical THR and RS (hip arthroplasty). The care pathways are similar in terms of pre and post-operative care and follow up.

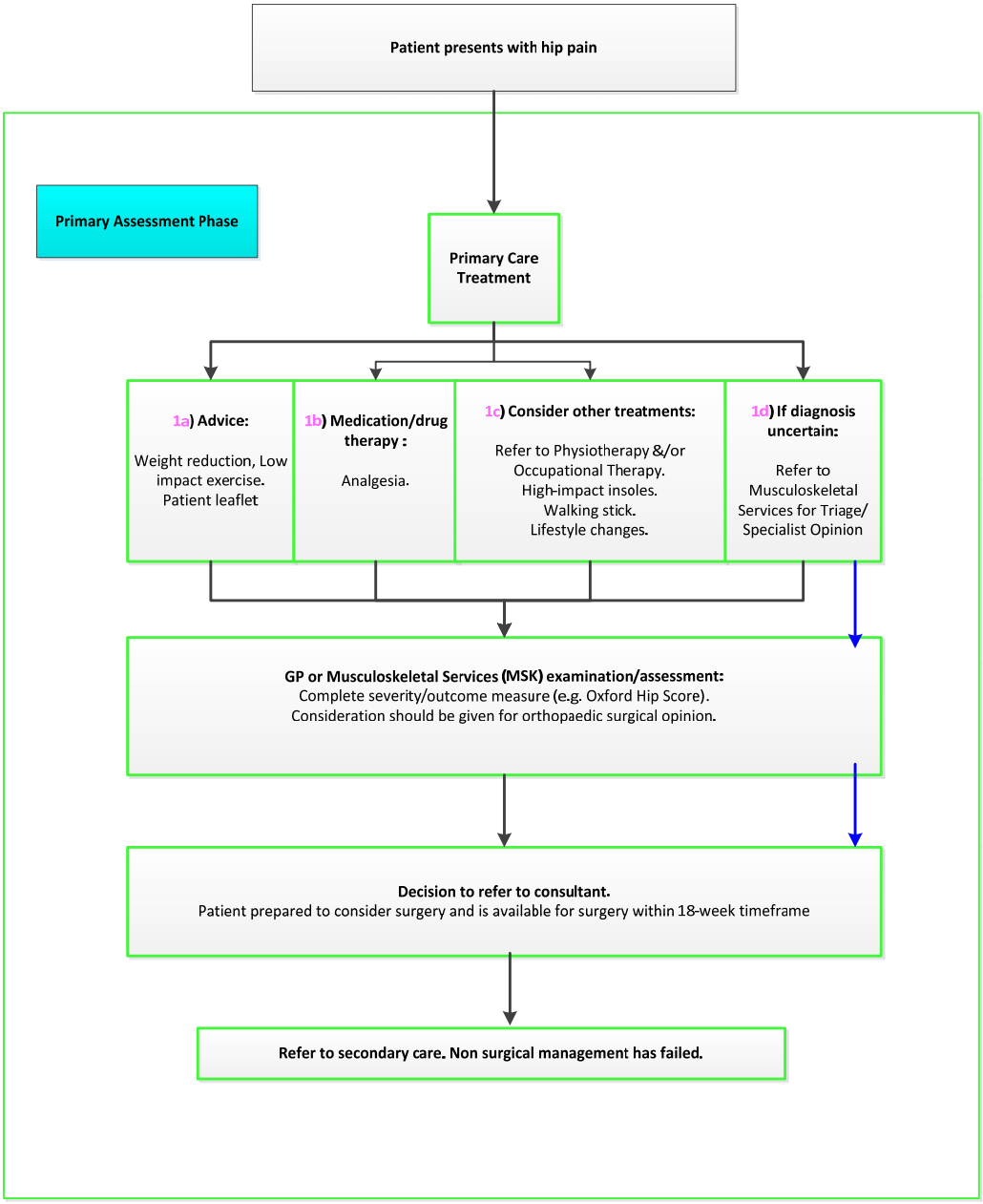


Figure 1. Example pathway for patient with arthritis in primary care

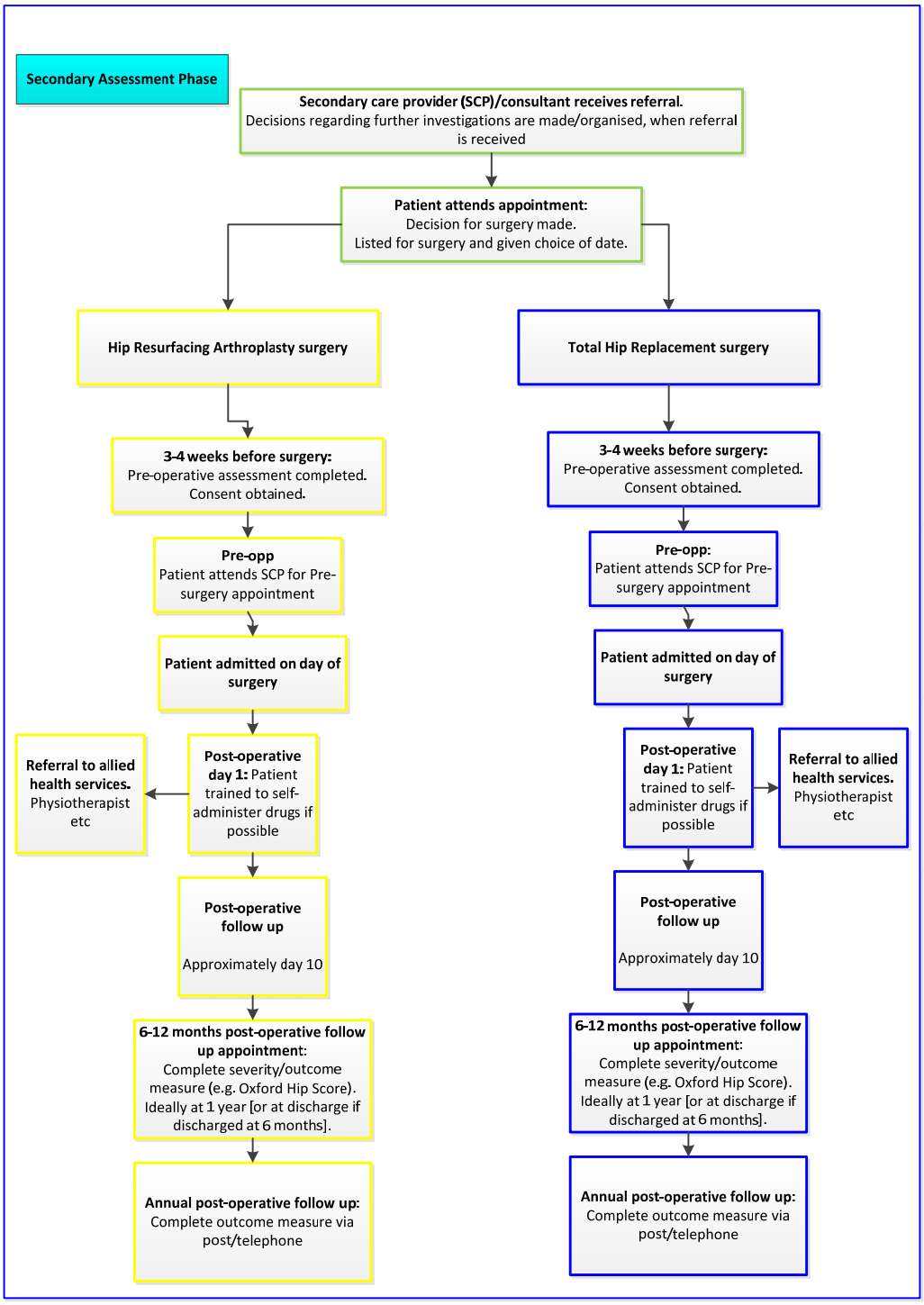


Figure 2. Example hip replacement care pathway in secondary care

3.2.1.2 Identification of different types of THR

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

a) Hip replacement with different fixation methods for implant components (cemented (Ce), cementless (CeL), hybrid (Hy) or reverse hybrid (RHy) prostheses)
b) Hip replacement with implant components (i.e., femoral stem, femoral head, acetabular cup) made from different materials (metal (M), ceramic (C), polyethylene (P))
c) Hip replacement with differing femoral head size

3.2.1.3 Hip replacement with different fixation methods

Hip replacement prostheses can be categorised by their fixation method (Figure 3) as (i), cemented (ii) cementless, (iii) reverse hybrid with a cemented cup and cementless stem (iv) or 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 National Joint Registry for England and Wales the percentage of cemented procedures did not change between 2009 and 2010 after being in steady decline since 2005 where the total per cent dropped from 77% in 2004 to 50% in 2010.[36]

Cementless prostheses reply on initial press-fit fixation followed by natural bone growth. They typically consist of four components, a femoral stem, femoral head, acetabular cup shell and 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 (HA)) which is compatible with bone growth and which helps to secure the liner in place. Hydroxyapatite is a mineral form of calcium apatite.[50] HA 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, while 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]

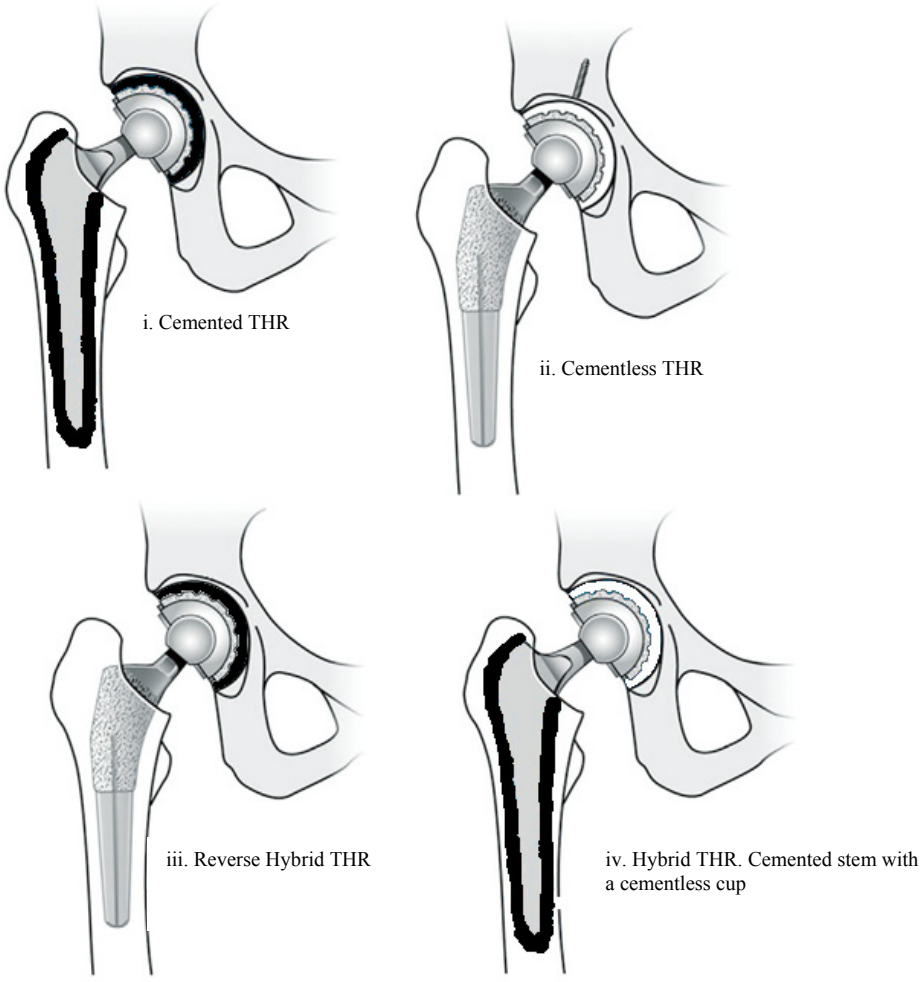


Figure 3. Overview of four different fixation options for the femoral stem and acetabular cup in THR

3.2.1.4 Hip replacement with components made from different materials

The combinations of prostheses 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. THR and RS articulation and fixation type combinations

	Femoral head (press fit)	**Fixation method**	**Femoral stem**	**Acetabular cup**	**Acetabular cup shell**	**Acetabular liner**
THR articulation type	Metal	Cemented (Ce)	Metal	Polyethylene		
	Metal		Metal	*Metal*		
	Ceramic		Metal	*Polyethylene*		
	Ceramic		Metal	*Ceramic*		
	Ceramic	Cementless (CeL)	Metal		Metal	Ceramic
	Metal		Metal		Metal	Polyethylene
	Metal		Metal		Metal	Metal
	Ceramic	Hybrid (Hy) 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 (RHy) cementless femoral stem and a cemented acetabular cup	Metal	Polyethylene		
	Metal		Metal	*Metal*		
	Ceramic		Metal	Polyethylene		
	Ceramic		Metal	*Ceramic*		
Resurfacing arthroplasty (RS)		**Fixation method**	**Femoral stem**	**Acetabular cup**		
		Cemented (Ce)	Metal	Metal		
		Cementless (CeL)	Metal	Metal		
		Hybrid (Hy)	Metal	Metal		

Ce=cemented, CeL=cementless, Hy=Hybrid, RHy=Reverse Hybrid
Italic = rarely used in clinical practice in England and Wales

The NJR report for 2011 stated the percent use of fixation type during 2010 and 2011. These can be seen in Table 2. The cemented fixation type was the most popular fixation method, and the metal-on-polyethylene 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 metal-on-polyethylene articulation combination was most popular (35.6%).

Table 2. Percent spread of fixation type and bearing/articulation surface for primary hip replacements reported in NJR 2011 (2010/11)

Articulation combination (femoral head material-on-cup material)	**Cemented (Ce) (n=132,511)**	**Cementless (CeL) (n=102,688)**	**Hybrid* (Hy) (n=43,933)**	**All (n=279,132)**
Other/unknown	2.9%	5.7%	3.8%	4.0%
Ceramic on ceramic (CoC)	1.8%	25.6%	15.1%	12.6%
Ceramic on polyethylene (CoP)	8.4%	14.2%	11.7%	11.0%
Metal on metal (MoM)	0.9%	18.9%	3.0%	7.9%
Metal on polyethylene (MoP)	86.1%	35.6%	66.5%	64.4%

*The NJR 2011 report did not distinguish between hybrid and reverse hybrid
Ce=cemented, CeL=cementless, M=Metal, P= Polyethylene, C=Ceramic, O=On, Hy=Hybrid

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 with the associated acronym that has been used for the remainder of this report.

Table 3. Combination of bearing/articulation surface and fixation method by frequency as reported in NJR 2010-11

Implant characteristics	**Acronym* for use in the report**
Cemented **p**oly cup **o**n **m**etal head (cemented stem)	CePoM
Cement**l**ess HA coated metal cup (**p**oly liner) **o**n **m**etal head (cementless stem)	CeLPoM
Cement**l**ess HA coated metal cup (**c**eramic liner) **o**n **c**eramic head (cementless stem)	CeLCoC
Hybrid cementless HA coated metal cup (**p**oly liner) **o**n **m**etal head (cemented stem)	HyPoM
Cementless non HA coated metal cup (**p**oly liner) **o**n **m**etal head (cementless stem)	CeLPoM (nonHA)
Cemented **p**olyethylene cup **o**n **c**eramic head (cemented stem)	CePoC
Hybrid cementless non HA coated metal cup (**p**oly liner) **o**n **m**etal head (cemented stem)	HyPoM (nonHA)

Ce=cemented, CeL=cementless, M=Metal, P= Polyethylene, C=Ceramic, O=On, Hy=Hybrid, HA=Hydroxyapatite
* Acronym order =(Fixation type), (Cup/liner material), (Femoral Head material)

Polyethylene-on-Metal (PoM) (cup material-on-femoral head material)

A metal ball with polyethylene cup (or polyethylene liner inside a metal cup) is the most common type of articulation combination (both cemented and cementless) and is one of the cheapest (Figure 4). The Charnley low-friction arthroplasty was the first widely accepted metal-on-polyethylene prosthesis to be used. It has a high reported implant survivorship at greater than 20-year follow-up (>80%) and 35-year follow-up (78%).[41] It also provides the baseline against which new prosthesis 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 due to its lower reported wear rates.[51,52]

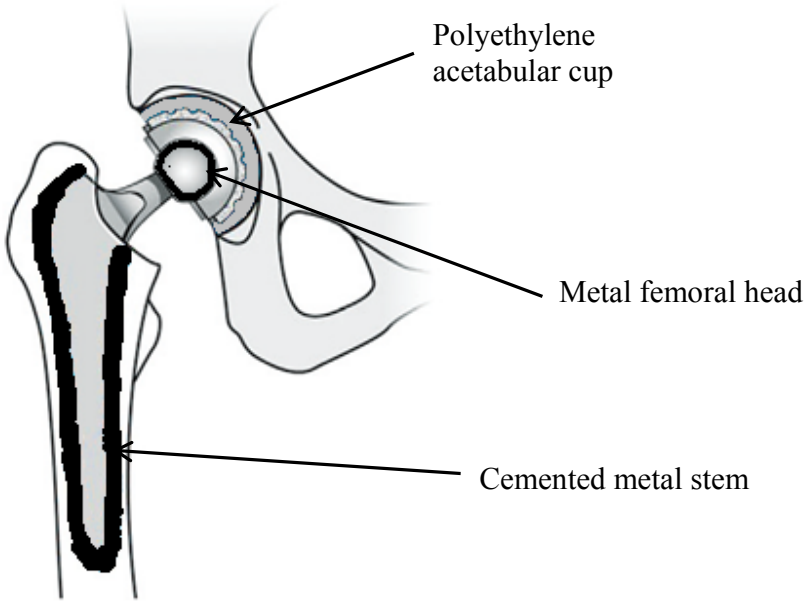


Figure 4. Cemented metal stem, metal femoral head and polyethylene acetabular cup

Polyethylene-on-Ceramic (PoC)

The PoC option combines polyethylene cup with a hard ceramic femoral head (Figure 5). This articulation type is reported to have a lower wear rate compared to the PoM bearing combination and is cheaper than the Ceramic-on-Ceramic (CoC) options. It is used more often with a cementless fixation (14.2%) than the cemented option (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.

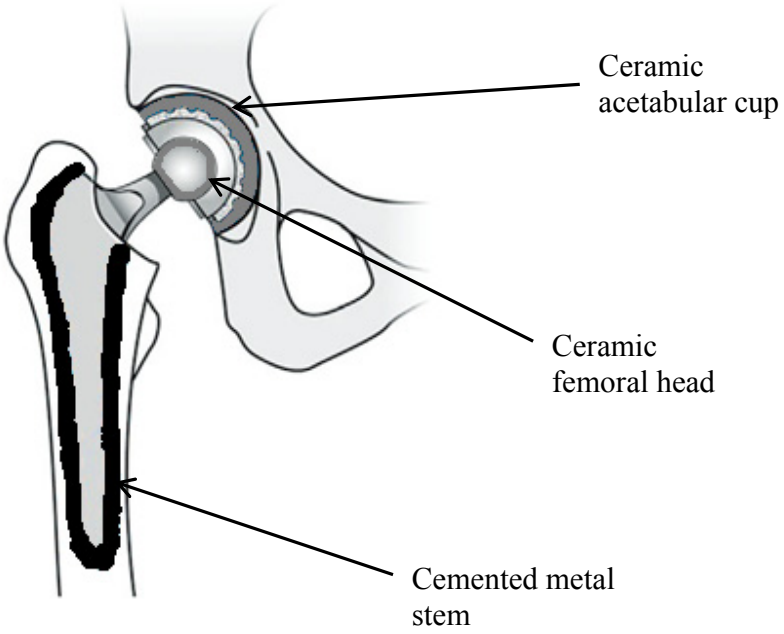


Figure 5. Cemented metal stem, ceramic femoral head and polyethylene acetabular cup

Ceramic-on-Ceramic

CoC articulation provide the hardest bearing surface combination and are generally the most expensive combination available[40] see Figure 6. These combinations have a lower reported wear rate than other options available to patients in England and Wales. CoC are mostly used without cement as shown in Table 2 (25.6%) compared with cemented (1.8%). Clinical advice suggested that the cementless ceramic cup is most common practice in England, cementing the ceramic cup is increasingly rare as demonstrated in the NJR data.

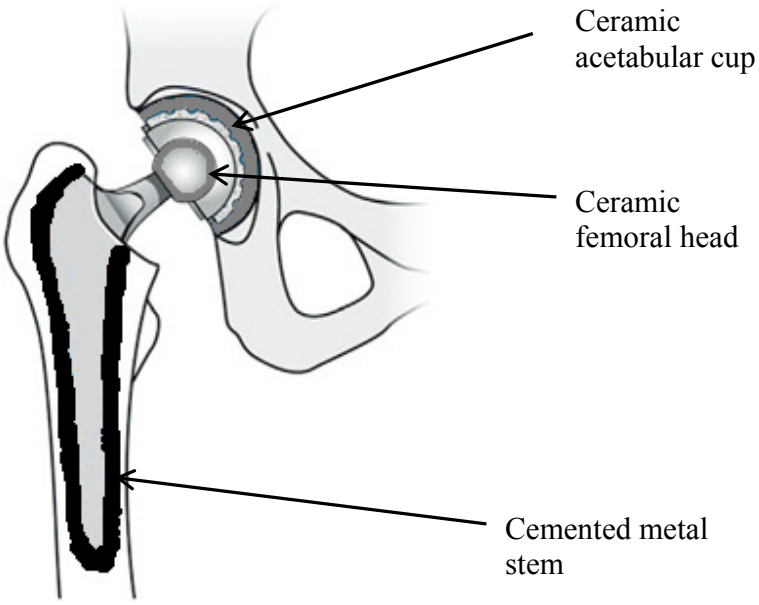


Figure 6. Cemented metal stem, ceramic femoral head and ceramic acetabular cup

Metal-on-Metal

MoM articulations provide a hard bearing surface, however due to their reportedly high revision rate they are no longer recommended by the MHRA for use in the UK (reference MDA/2012/036) (see Figure 7).

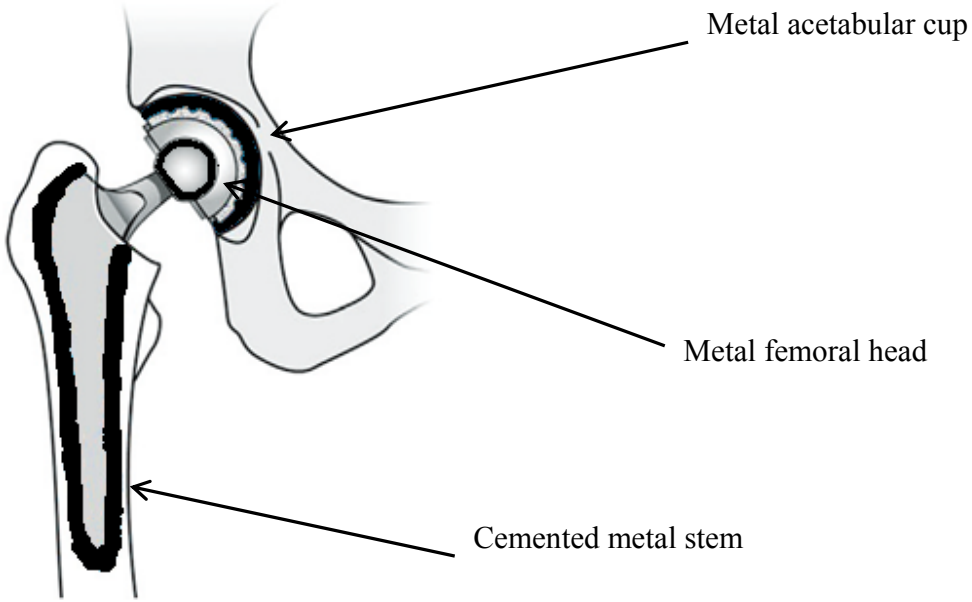


Figure 7. Cemented metal stem, metal femoral head and metal acetabular cup

The MHRA recommendations apply to four groups of MoM replacements:

1) MoM hip resurfacing implants
2) MoM total hip replacements with a head diameter <36mm
3) MoM total hip replacements with a head diameter ≥36mm
4) DePuy ASR™ hip replacements comprising:
    - ASR™ acetabular cups for hip resurfacing arthroplasty or total hip replacement
    - ASR™ surface replacement heads for hip resurfacing arthroplasty
    - ASR™ XL femoral heads for total hip replacement.

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

Head size is important because in simulator tests larger head sizes give lower wear due to the boundary lubrication regime becoming more favorable.[54] Therefore, implantation of large diameter MoM 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 particularly appropriate for younger, more active patients.

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

This prompted recent analysis of NJR for England and Wales data regarding 402,051 hip replacements to assess whether MoM 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.

Results revealed that in THR MoM articulations failed at higher rates than other bearings. For example, five year revision rates in younger women were 6.1% (5.2-7.2) for 46mm MoM compared with 1.6% (1.3-2.1) for 28mm MoP. This effect was found, even though the ASR data had been removed before analysis (these DePuy ASR™ articulations had already been removed from the market). Thus, it is a problem with all MoM 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 than for CoC articulations). The authors suggested a number of potential reasons for the finding that the larger head metal heads fail earlier, such as failure to achieve optimum lubrication or trunion (post which inserts into head) wear [55] resulting in metal debris leading to local soft tissue reactions [57] or early loosening due to increased transmitted torque from the larger head. These authors therefore recommended that MoM replacements not be performed due to poor implant survival. They also suggest that all patients with existing MoM THR undergo at least annual review with both clinical and radiological examination for the duration of the longevity of the implant.

Furthermore, there are the potential dangers of exposure to metals such as chrome and cobalt. Metal alloys used in MoM bearings degrade through wear, from corrosion, or by a combination of the two.[58] Consequently, they produce a vast number of nanometer to submicrometer sized metal particles which cumulatively present a large surface area for corrosion.[59] This is also relevant to the metal-on-

polyethylene 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 has 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 due to osteolysis induced by some immunological reaction involving hypersensitivity[62] and local pseudotumours (soft tissue masses relating to the joint) which 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 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 by either direct action, causing DNA breaks through attacks on free radicals or by 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 certain malignancies in MoM articulation patients.[69]

The US Food and Drug Administration (FDA), the UK MHRA, and the British Orthopaedic Association have released statements of concern about metal-on-metal articulations. The MHRA recommendation states that patients with MoM bearings and a painful hip joint should have yearly measurements of serum 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.[47] Although the use of MoM bearing surfaces has consequently declined in England and Wales, 7.9% of all procedures in 2010-2011 (see Table 2), data suggests that they are still being extensively used in other countries. For example the USA; 35% of articulations were MoM in 2009.[70]

3.2.1.5 Hip replacement with differing femoral head size

Research has suggested that differing femoral head sizes lead to variation in the rate of revision. Smith 2012 and colleagues reported that the use of larger head sizes (greater than 36 mm in diameter) improves stability and range of motion compared with the smaller head diameters that are used with other bearing surfaces.[16] Use of large diameter femoral heads increases the distance that the head must travel before dislocation, without decreasing hip range of motion, and thus increasing stability.[41]

3.2.2 Summary of hip resurfacing arthroplasty (RS)

Hip RS arthroplasty has been developed as a surgical alternative THR. It is reported to be an option predominantly suited to younger, active, male patients.[46] The procedure consists of placing a cobalt-chrome metal cap, over the head of the femur while a matching metal cup (similar to 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.

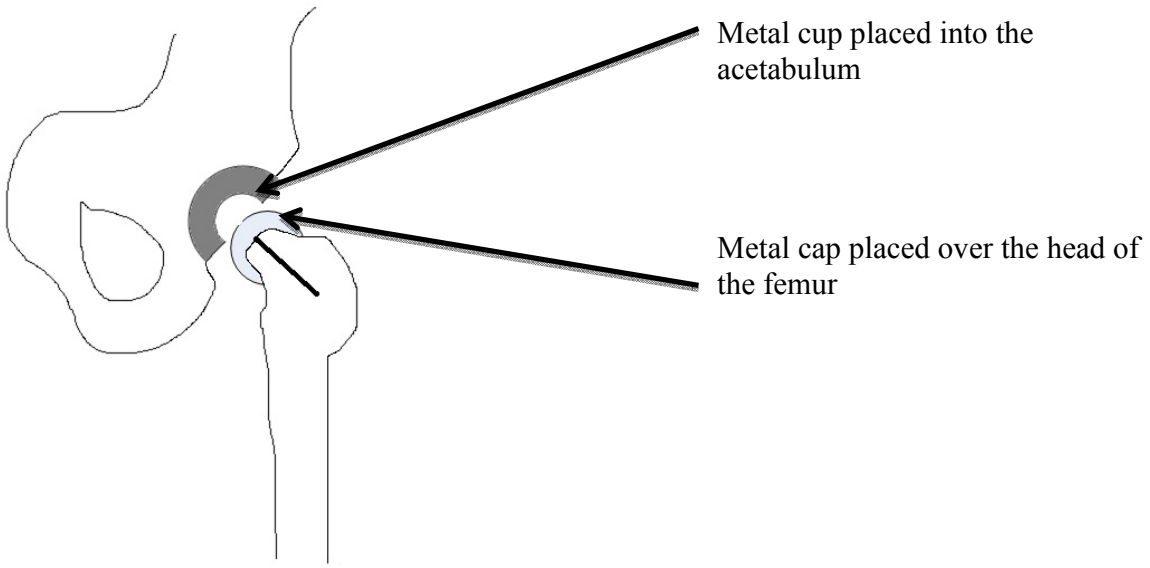


Figure 8. Diagrammatic representation of a hip resurfacing arthroplasty (RS)

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

The FDA have produced patient selection criteria for metal on metal hip RS arthroplasty. 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.[71]

Johnson and colleagues reported 100% implant survivorship at five-years follow-up in 93 patients having RS identified using narrow selection criteria.[72] The selection criteria included avoiding RS in patients with large femoral head or neck cysts, ensuring proper seating of the femoral component band ensuring an optimal thickness of the cement mantle. They suggested that the best results were achieved in male patients younger than 50 years, with a primary diagnosis of OA, and a native femoral head greater than 50 mm in diameter.[72] 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 which must be negotiated for surgeons inexperienced with the procedure.[41]

Over the last year in England and Wales there has been a significant decrease in the percentage of RS procedures and in the percentage of procedures where a large head is used with a RS cup taking place.[36] This is thought to be due to the withdrawal of the ASR-DePuy RS device from the market following the identification of higher than expected revision rates for this product.

3.2.3 Failure of hip replacement

The failure of hip replacement may occur due to peri-and/or post-operative complications such as implant instability, dislocation, aseptic loosening, osteolysis, implant fracture, and infection.

3.2.3.1 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. Prevalence of dislocation ranges between 0.3% and 10% for primary THR, and as 28% for revision THR.[73-75]

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. For example, 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.[76,77] However,

age, previous fracture, surgical volume, surgical approach, component sizing and polyethylene wear are also contributory factors to revision due to instability and dislocation.[78-80 81]

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.

3.2.3.2 Aseptic loosening and osteolysis

Aseptic loosening is a common cause of failure of THR. It arises because of osteoclast-mediated bone re -absorption at the bone-implant interface, which can lead to loosening, implant migration, implant failure, and periprosthetic fracture.[82] Osteolysis is one of the most common complications after THR which may lead to implant failure. It is initiated due to inflammatory process against polyethylene particular debris. Component malpositioning is a major cause of severe wear and osteolysis, but it is also affected by activity level and material and component design.[83]

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 less than 5% at follow-up.[84]

3.2.3.3 Periprosthetic fracture

Periprosthetic fracture is a major complication after THR and is associated with increased morbidity and mortality. Risk factors for periprosthetic fracture include prior revision surgery, component malalignment, age, osteoporosis, previous fracture, and minor trauma.[85,86]

Treatment for most periprosthetic fractures is usually surgical. Options depend on the fracture pattern but include open reduction and internal fixation with or without cortical strut allografts, longer femoral stems, or changes increases in the setting of acetabular fractures, or tumour prostheses.[87,88]

3.2.3.4 Infection

Infection of a THR prosthesis is associated with greatly increased morbidity, mortality, and use of healthcare 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 due to 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.[89,90]

3.2.4 Revision of hip arthroplasty

Recent data demonstrated that seven year revision rates were lower for cemented (3.0%) than for hybrid (3.8%) or cementless prostheses (4.6%).[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 the implant has failed, patients will go to 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 see Table 4.

Table 4. Revision procedures by type and year as published NJR data

Procedure by year and type	2006/7	2007/8	2008/9	2009/10	2010/11
Hip primary	58,445	66,556	69,681	70,669	77,800
Hip revision	6,198 (9.6%)	6,725 (9.2%)	7,345 (9.5%)	8,285 (10.4%)	9,200 (10.6%)
Total	64,643	73,281	77,026	78,954	87,000

This shows a rise in the number and proportions of operations which 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 five years. This may be due to the recipients of the replacements living longer and thus outliving their THR, or possibly due to more stringent follow-up. At NHS hospitals, revision procedures account for a higher percentage of total procedures (13%) than at any other type of provider, 84% of all revision procedures in 2010/2011 were performed in the NHS.[36]

3.2.4.1 Clinical follow up

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

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.[91] Earlier detection of lesions, such as aseptic lymphocyte dominated vasculitis, may place the implant at risk and has been reported to be more cost-effective than is assessment of patients by the time pain or loss of function are present.[92]

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-(SF)12/36, Euro-Qol EQ-5D).[27]

3.3 Current usage in the NHS

Information taken from the NJR for England and Wales 2011.[36]

3.3.1 General statistics

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

3.3.2 Hip replacement surgery

Where the operations took place: (2010/11 data)[36]

- England: 83,014 (95%)
- Wales: 4,024

There are four types of organisation in England carrying out hip replacement surgery: (Note: there are no NHS treatment centres or independent sector treatment centres in Wales). Please see Table 5.

Table 5. Percentage of procedures by organisation type reported in the NJR 2010/11[36]

Organisation type	Percentage of procedures in 2010/11
NHS hospitals	67%
NHS treatment centres	3%

Independent sector hospitals	26%
Independent sector treatment centres (ISTCs)	5%

There have been no major changes in these proportions over the last five years, although there has been a constant, very slight increase in the proportion 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 for each type of provider in 2010/11 is within two percentage points of the figure from 2006/7. Ninety three per cent of patients at Independent Hospitals and ISTCs reported patients as 'fit and healthy' or with 'mild' disease (ASA grading system), compared with only 80% at NHS centres.

3.3.3 Type of procedure

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

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

Table 6. Percentage of intervention by fixation method across NHS hospitals and treatment centres reported in the NJR 2010/11[36]

Procedure type	Overall (68, 907 treatments)	NHS hospitals (44,054 treatments)	NHS treatment centres (2,075 treatments)
Cemented (Ce)	36%	38%	25%
Cementless (CeL)	43%	42%	66%
Hybrid (Hy)	3%	17%	4%
Resurfacing (RS)	2%	3%	4%

3.4 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.

OA 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 ones. OA is by far the most common arthritis of the hip, and is diagnosed clinically together with imaging. There are difficulties in estimating disease burden of OA due to 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 and over will be affected at some point. Estimates of 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 ages 40 and 60. This is especially important in light of the ageing population since OA and RA mostly affect elderly people with comorbidities. Although the natural history of OA varies between affected joint, 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 due to direct costs to the healthcare system, community and social services and indirect costs due to lost productivity and early mortality. In the present economic climate where healthcare 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 THRs, are usually considered for patients with symptoms unmanageable via conservative management. The surgical interventions are believed to be cost-effective interventions which maximise cost per quality life year (QALY) gained. Patient selection criteria, amount spent and outcomes for hip replacement surgery vary across geographical population, hospital and surgeon. The NCC-CC and NICE have developed guidelines to assist clinicians with making clinical choices regarding whether a patient requires a hip replacement, however there still exist inconsistencies in surgeries offered at different NHS centres.

THR 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 arthroplasty. Failure of the articulations and need for revision surgery are an important

consideration, especially considering the growing number 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.

4 DEFINITION OF THE DECISION PROBLEM

4.1 Decision problem

This report aims to evaluate the clinical 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:

i. Suitable for both procedures, what is the clinical effectiveness and cost-effectiveness of different types of elective primary THR compared to primary hip RS arthroplasty?
ii. Not suitable for hip RS, what is the clinical effectiveness and cost-effectiveness of different types of primary THR compared with each other

4.2 Overall aims and objectives

1) To undertake a systematic review of the clinical and cost-effectiveness of the following:

A. Different types of primary THR compared with RS for people in whom both procedures are suitable
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 develop the cost-effectiveness and cost-utility models published in the 2002 HTA (Technology Appraisal No. 44, 2002) further using updated National Joint Registry data and model inputs where available.

3) To report on findings and make recommendations for future research.

Table 7. PICO table

PICO	Final scope issued by NICE (17.01.13)	Decision problem addressed in the assessment report	Comments
Population	People with pain or disability resulting from arthritis of the hip for which 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	'end stage' description agreed at subsequent NICE meeting
Intervention	1. Primary total hip replacement 2. Primary hip resurfacing arthroplasty	1. Elective primary total hip replacement 2. Primary hip RS arthroplasty	Elective added to ensure that treatment of trauma patients are excluded as specified
Comparators	Different types of primary total hip replacement and hip resurfacing arthroplasty will be compared with each other for people in whom both procedures are suitable Different types of primary total hip replacement will be compared with each other for people in whom hip resurfacing arthroplasty 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 size • Prostheses with differing revision rates	Different types of primary THR and hip RS arthroplasty for people in whom both procedures are suitable Different types of primary THR compared with each other for people in whom hip RS arthroplasty is not suitable	More specific detail in scope of 'different types'
Outcomes	The outcome measures to be considered include: • Functional result • Pain • Bone conservation • Revision rates • Radiosteriometric analysis to assess prosthesismovement • Dislocation rates • Adverse effects of treatment (peri- and postprocedural),including degradation products were appropriate • Health-related quality of life • 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 result, dislocation rates, health related quality of life and mortality Adverse events include peri- and post-procedural complications (e.g. infection, nerve palsy, dislocation rates, femoral neck fracture, metallosis, muscle weakness) and metal and other degradation products	More specific detail in report on adverse events
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per éequality-adjusted life year	Cost-effectiveness outcomes include mean difference in costs and clinical effectiveness measures or utility measures; incremental cost-effectiveness ratio (ICER),	

	The reference case stipulates that the time horizon for estimating clinical 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 an NHS and Personal Social Services perspective	uncertainty measures, ceiling willingness-to-pay ratios, and probabilities from CEACs	
Different types of THR to be considered	If the evidence allows subgroups based on activity levels will be compared. Guidance will only be issued in accordance with Conformité Européene marking 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 size	See above comments

5 JOINT REGISTRIES

5.1 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 i.e., 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 worldwide include: the UK, Canada, Australia, New Zealand, Sweden, Italy, Norway, and Denmark (among others) see Table 8. 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. Joint (hip) replacement registries available worldwide
*more than 1000 entries

Name	Country	Year established	Lifetime reporting	Most recent report	Data collected
National Joint Registry	England and Wales	2003	10 years	2011 Surgical data to 31st 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 - Revision procedures (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 gain (EQ-5D index gain after one year) - Patient satisfaction after one year - Short-term complications after two years - Ten-year implant survival (95%) - Compliance (98.5%)
Australian Orthopaedic Association established the National Joint Replacement Registry	Australia	1999	13	2012	Reports outcome variables across all states: - Ten-year implant survival (95%) - RS reported to be 1.6% procedures - Compliance (93.9%)

5.1.1 Australian Orthopeadic Association National Joint Registry

The Australian Orthopaedic Association established the National Joint Replacement Registry (AOANJRR) in 1993. 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.[93] The register was expanded to include other joint replacements in November 2007 with all hospitals undertaking joint replacement in Australia approving participation of the additional data collection. More than 37,000 hip replacements were undertaken in Australia in 2012. The total figure has been steadily increasing since 1999.[93]

The most recent report from the AOANJRR discussed the large increase in revision hip procedures in Australia.[93] In 2010, revision procedures represented 11.3% of all hip replacements but by 2011, this had increased to 12.5%. The authors associated this increase with the DePuyASR hip (discontinued metal on metal hip replacement) and its reported problems. The use of primary total RS hip replacement 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 prostheses combinations was reported. In 2010 there were 330 combinations being used in Australia. This had reduced to 97 in 2011.

5.1.2 The Swedish Hip Arthroplasty Register

The Swedish Hip Arthroplasty Register (SHAR) is entering its 33rd year of activity.[94] Nation 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 patients' data 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 at the one-year follow-up is just over 90%.

All units in Sweden (78 hospitals), public and private, which carry out total hip arthroplasty are included in the Registry. The registry's aim is to identify predictors for both good and poor outcome.[94] Sweden has the world's highest reported 10-year implant survival for total hip arthroplasties in international comparisons. At county council level there are no large and significant differences which are detectable at unit level. The 10-year survival of the most common implants was over 95% in 2010.[94] The 2010 report stated that the potential for improvement lies chiefly among certain patient groups. Even though Sweden has the lowest reported frequency of revision, there are still clearly defined problem areas which can be influenced with systematic local analyses and subsequent work for improvement.

5.1.3 National Joint Registry for England and Wales

The NJR for England and Wales aims to improve patient safety and clinical outcomes by providing information to all those involved in the management and delivery of joint replacement surgery, and to patients. This is achieved by collecting data in order 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 1st April 2003. The most recent report was from the period 1st April 2010 to 31st March 2011 and also included statistics on joint replacement activity and a survivorship analysis of hip replacement surgery using data from 1st April 2003 to 31st 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 1st April 2003 to 31st 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
- Patient consent rate of 90.4%
- Activity and outcomes data was reported 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 pre-operative PROMs questionnaire in accordance with the relevant guidance. Post-operative 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. The NJR are currently working to extend their own study of the follow-up PROMs to 12 months. This will allow for investigation of population level quality of life reporting after hip replacement.[36]

5.2 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 cost associated with development and maintenance of national joint registries varies, registries are considered a beneficial medical development due to their ability to detect poorly performing implants at a national level.

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

6 ASSESSMENT OF EVIDENCE

6.1 Clinical effectiveness methods

A protocol was developed and approved by the NICE (www.nice.org.uk/nicemedia/live/13690/62831/62831.pdf; see Appendix 1). General principles were applied as recommended by the NHS Centre for Reviews and Dissemination (CRD).[95]

6.1.1 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 2). An iterative procedure was used to develop these strategies with input from clinical advisors and previous HTA reports (e.g., Vale et al., 2002;[19] deVerteuil et al., 2008[11]). The strategies have been designed to capture generic terms for arthritis, THR and RS.

6.1.2 Search strategies

Final searches were undertaken in November and December 2012 (see Appendix 2) 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 randomised controlled trials (RCTs) and systematic reviews; additional searches were undertaken to capture literature relating to costs, resources 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; The Cochrane Library (specifically Cochrane Database of Systematic Reviews, CENTRAL, DARE, NHS EED, HTA database), Current Controlled Trials; ClinicalTrials.gov; 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
- Joint Replacement Instrumentation (JRI)
- Implantcast
- Implants International
- Lima WG Healthcare
- Mathys Orthopaedics
- Medacta UK
- Othodynamics
- Peter Brehm
- SERF dedienne santé
- Smith & Nephew
- Stanmore Implants Worldwide
- Stryker
- Symbios SA
- Waldemar Link
- Wright Medical UK
- Zimmer

Grey literature searches were undertaken using Google 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
- 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.

6.1.3 Inclusion criteria

Study design:

- Randomised controlled trials
- Systematic reviews
- Meta-analyses

Given the wide scope and large amount of identified evidence, we limited our inclusion to studies published since 2008 with a sample size of 100 participants or more.

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 arthroplasty

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/post-procedural complications (i.e., implant dislocation, infection, osteolysis, aseptic loosening, femoral fracture, and deep vein thrombosis).

6.1.4 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

6.1.5 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 PRISMA diagram.[96]

6.1.6 Quality assessment strategy

Two reviewers independently assessed the risk of bias of individual studies using validated tools (see Appendix 3).[97,98] Any disagreements between the two reviewers were resolved by a third reviewer through a discussion.

RCTs were assessed using the Cochrane Collaboration Risk of Bias tool (ROB)[97] 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 pre-specified 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 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 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 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.

Methodological quality of included systematic reviews was assessed with the AMSTAR tool[98] which covers the following domains: a) research question, b) inclusion/exclusion criteria, c) search strategy (at least two major electronic databases), d) data extraction by independent reviewers, e) assessment of risk of bias by independent reviewers, f) consideration of risk of bias in the analysis, g) exploration of heterogeneity, and h) 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).

6.1.7 Grading overall quality of clinical effectiveness evidence

The overall quality of evidence for each pre-selected (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 (http://www.gradeworkinggroup.org).

The GRADE approach[99] indicates levels of confidence in the observed treatment effect estimate(s), which is categorized 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: a) summary risk of bias across studies per gradable outcome (internal validity across studies; study limitations), b) consistency of results (heterogeneity), c) directness of the evidence (applicability of the results; indirect treatment comparisons), d) precision of the results (the width of 95% CI around the estimate), and e) publication/reporting bias (detection of asymmetry in the funnel plot; selective outcome reporting). The definitions and explanation

of the grading levels and the grading process across the five domains are presented in Sections 6.4.3 and 6.4.6 (see Table 35 and Table 43).

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

6.1.8 Data extraction strategy

The relevant data were extracted from included studies independently by one reviewer using a data extraction form informed by the NHS Centre for Reviews and Dissemination (CRD)[100] The extracted data was 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 4 and 5, respectively). The extracted information included the following:

- Study characteristics (i.e., author's name, country, design, study setting, sample size, funding source, duration of follow-up, information relevant to risk of bias assessment such as generation of randomization, allocation concealment, blinding, completeness of outcome ascertainment, patient withdrawals/attrition for randomised trials; for observational studies and non-randomised trials, information on potential confounding was additionally ascertained)
- Patient baseline characteristics (i.e., inclusion/exclusion criteria, number of enrolled/analysed participants, age, race, gender, body mass index, underlying conditions, concomitant conditions, co-interventions, disability, activity levels, function, pain intensity, and quality of life, and disease-specific measures such as the Oxford Hip Score,[30] and Harris Hip Score[31])
- Experimental treatment characteristics (e.g., type - THR, RS; training/experience of the operator, post-operative rehabilitation staff; method of fixation – cemented, cementless, hybrid; bearing surface material – metal-on-metal, ceramic-on-ceramic; metal-on-polyethylene, femoral head size; the name/brand and country of manufacturer; post-operative rehabilitation)
- Outcome characteristics (e.g., definition; timing of measurement; scale of measurement - dichotomous, continuous; measures of association – mean difference, risk ratio, odds ratio, hazard ratio). Statistical test results and measures of variability were also extracted (standard deviation, 95% CIs, standard error, p-values)

Any additional relevant information found in multiple publications of included studies was also extracted. For studies of clinical effectiveness where summary measures and 95% CIs for the association between

the treatments were not reported, mean differences with 95% CIs were calculated, if data allowed (t-tests for independent samples and using continuous outcomes and risk ratios for dichotomous outcomes). No risk ratios and 95% CIs were estimated for individual studies which observed zero events in one or both treatment arms. The 95% CIs and standard errors were used to derive standard deviations or vice versa. All calculated parameters were entered into the data extraction sheets.

6.1.9 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

6.1.10 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 mean difference (MD) for continuous outcomes and risk ratios (RR) for binary outcomes (except for rare events) of individual studies were pooled using a DerSimonian and Laird random-effects model (DerSimonian & Laird, 1986).[101] 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 RR using a Mantel-Haenszel fixed-effects model. Dichotomous outcomes for studies with very low event rates (≤ 5.0%) or zero events in one of the treatment arms were pooled as odds ratio (OR) using a Peto fixed-effects model.[102]

Trials were not pooled if the mean and/or standard deviation 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 statistics. The presence of heterogeneity was judged according to pre-

determined levels of statistical significance (Chi-square $p < 0.10$ and/or $I^2 > 50\%$). Statistical pooling was performed using the Cochrane Collaboration software package Review Manager version 5.2.

6.1.11 Publication bias

Extent of publication bias, given a sufficient number of data points, was planned to be examined by visual inspection of funnel plots with respect to plot asymmetry as well as using linear regression tests.[103]

6.1.12 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.

6.1.13 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 'difference' or there is 'no difference') or inconclusive evidence (indeterminate results due to statistical uncertainty, statistical heterogeneity/inconsistency in treatment effects, and/or incomplete information). This conclusion was based on several factors taken 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 (MCID: 0.074) from previous empirical research evidence.[104-106]

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 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 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 CIs sufficiently wide to include the MCID or any large effect size values. (Since 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 insufficient sample size, short follow-up periods, and/or low event counts, leading to an inadequate study power and increased chance of type II error).

The results were also considered inconclusive if there was partially missing data for continuous outcomes (e.g., reporting treatment arm-specific means without standard deviations; reporting only p values for the between-treatment difference) or zero events for binary outcomes in both treatment arms. Evidence consisting of studies showing inconsistent results, i.e., 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 synthesis), those reporting inappropriate pooling methods (e.g., indirect naïve comparison of single group cohorts; pooling of studies of different design), or those reporting inconsistent summary findings were also considered inconclusive.

6.1.14 Industry submissions regarding effectiveness of treatments

The included clinical effectiveness evidence was compared to the evidence submitted by industry. These industry submissions will be discussed in Section 10.5 and Appendix 6.

6.2 Clinical effectiveness results

6.2.1.1 Search results

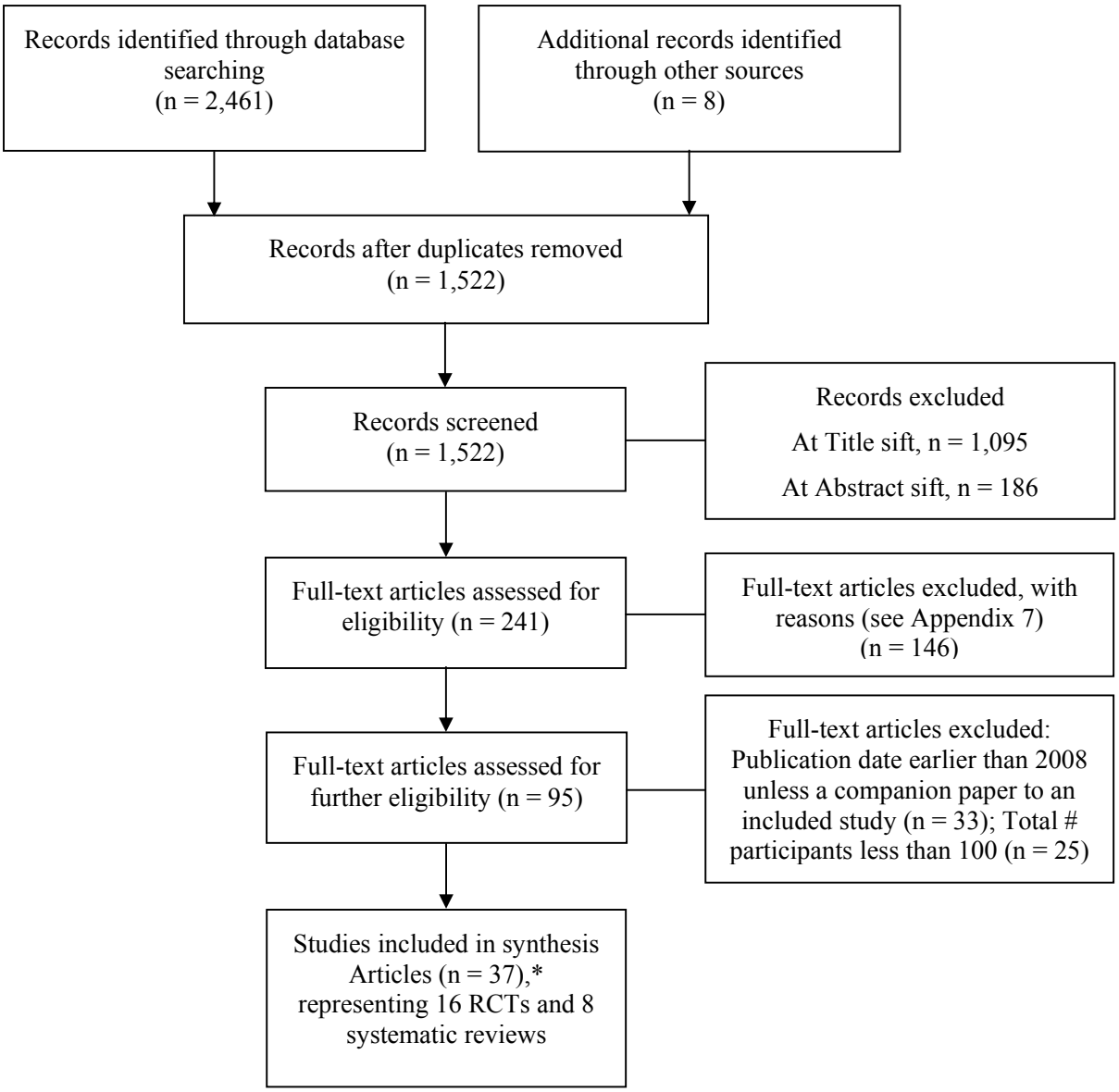
A total of 2,469 records were identified through our searches of different sources. The removal of duplicates left 1,522 records to be screened. Of these, 1,281 records were excluded as irrelevant at title and abstract screening, leaving 241 potentially relevant records. Of the 241 full text records 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 (less than 100 participants). The remaining 37 records were included in the review.[104,107,108,108-141]

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

A list of records excluded at full screen with reasons for exclusion is provided in Appendix 7. The main reasons for exclusion were the comparison of different surgical/operative approaches, (n = 42),[11,142-182]

publications before 2008 (unless a companion paper to an included study) (n = 33)[19,39,183-213] and total number of participants less than 100 (n = 25).[81,214-237]

A separate search (December 2012) of the Clinical Trials.gov, Current Controlled Trials, UKCRN Portfolio, and Health Services Research Projects in Progress (HSRProj) databases 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,[127] had already been identified from the original searched database (see Appendix 8). The identified clinical trials were considered potentially relevant based on the available information. All were either on-going, completed since 2009, or their status was unknown.



*A further 20 on-going clinical trials were identified

Figure 9. PRISMA 2009 flow diagram of clinical effectiveness

The included 37 records represent 16 RCTs[104,107-133,238] and eight systematic reviews.[134-141]
Six of the 16 RCTs were represented by multiple publications:

1. Bjorgul 2010[107] and Bjorgul 2010[108]
2. Engh 2012[110] and Engh 2006[111]
3. Capello 2008,[112] D'Antonio 2005,[113] D'Antonio 2003,[114] and Mesko 2011[115]
4. Corten 2011,[116] Laupacis 2002,[117] Bourne 2010,[118] and Corten 2011[119]
5. Costa 2012[127] and Achten 2010[104]
6. Vendittoli 2010,[129] Vendittoli 2006,[130] Girard 2006,[131] Rama 2009,[132] and Vendittoli 2006[133]

The six RCTs mentioned above are cited as follows: Bjorgul 2010,[107] Engh 2012,[110] Capello 2008,[112] Corten 2011,[116] Costa 2012,[127] and Vendittoli 2010.[129] Thirteen RCTs[107,109,110,112,116,120-126,238] and five systematic reviews [134-138] comparing different types of primary THRs and three RCTs [127-129] and three systematic reviews [139-141] comparing primary THR to RS arthroplasty were finally included in the current review.

In the following sections we will begin by reporting the findings for THR vs THR and then findings for THR vs RS .

6.2.2 Comparison of total hip replacement (THR vs. THR)

6.2.2.1 Study and participant characteristics

RCTs

The study and participant characteristics of the 13 included RCTs[107,109,110,112,116,120-126,238] are summarised in Table 9. More details can be found in Appendices 4 and 5. Briefly, four RCTs were conducted in the USA,[110,112,122,124] one in the UK,[109] one in Australia,[120] two in Norway[107,123] two in South Korea,[125,126] and three in Canada.[108,116,121] A total of 3,175 participants were randomised across the 13 RCTs ranging from 100[121,125,238] to 557 participants.[120] The mean age of participants across the RCTs ranged from 45 [126] to 72 years.[120,238] The proportion of women across the studies ranged from 24%[126] to 73%.[107] The length of follow-up of the studies ranged from three months[116] to 20 years.[116,126] Proportion of participants diagnosed with primary osteoarthritis was reported for nine studies[107,109,110,112,120,121,124-126] and ranged from 14%[126] to 96%.[120]

Table 9. Overall study characteristics across 13 RCTs comparing THRs

Study Characteristic	Frequency
Geographical region	UK (n = 1); Australia (n = 1); Norway (n = 2); South Korea (n = 2); Canada (n = 3); USA (n=4)
Total number of randomised participants	3,175 (range: 100 - 557)
Mean age (in years)	Range: 45 - 72
Female participants (%)	Range: 24% - 73%
Length of follow-up	Range: 3 months - 20 years
Diagnosis of primary osteoarthritis (%)	Range: 14% - 96%

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,[124] design,[112,125] bearing surface,[110,112,121-123,238] fixation method,[107,109,116,126] and component size.[120] Table 10 shows the distribution of RCTs across the THR comparison categories.

Table 10. Distribution of 13 RCTs according to basis of THR comparison

Basis of comparison	Study ID
1. Cup fixation	Bjorgul 2010[107]
	Angadi 2012[109]
2. Cup liner bearing surface	McCalden 2009[238]
	Engh 2012[110]
3. Cup shell design	Capello 2008[112]
4. Cup/stem fixation	Corten 2011[116]
5. Femoral head size	Howie 2012[120]
6. Femoral head bearing	Lewis 2008[121]
7. Femoral head on cup liner bearing surface	Amanatullah 2011[122]
	Capello 2008[112]
	Kadar 2011[123]
8. Stem composition	Healy 2009[124]
9. Stem design	Kim 2011[125]
10. Stem fixation	Kim 2011[126]

Reported outcomes across the 13 RCTs varied. Most RCTs reported HHS,[107,109,110,112,116,121-126,238] risk of revision,[109,110,112,116,120-122,124-126] The follow-up of outcome assessments ranged from three months[116] to 20 years.[116,126] Outcomes reported in the included studies can be found in Appendix 9. A summary of the functional/clinical and quality of life measures/tools is provided in Appendix 10.

Systematic reviews

The five included systematic reviews evaluated RCTs and non-RCTs of clinical effectiveness of THRs (see Appendix 4)[134-138]. The primary focus of these systematic reviews was the comparison of effects of

different cup fixation methods (cemented vs. cementless)[134-136] and materials used for implant articulations[137,138] on post-operative clinical/functional scores (HHS, OHS),[134,135,137] risk for revision rate.[135,136] Searches were undertaken between July 2007[138] and June 2011.[136] Further details on specific outcomes reported in the included systematic reviews can be found in Appendix 9.

6.2.2.2 Risk of bias and methodological quality

Risk of bias in RCTs

Risk of bias assessment for the 13 included RCTs comparing different types of THR are presented in risk of bias tables (Appendix 3), the summary table (

Table 11) and risk of bias graph (Figure 10). Overall, four[109,116,120,125] of the 13 RCTs reported an adequate method for random sequence generation and eight[107,109,116,120-123,126] reported adequate treatment allocation concealment (low risk of bias). A greater proportion of the RCTs were rated as having low risk of performance and detection bias for objective (e.g., mortality, dislocation) vs. subjective (e.g., patient-administered functional scores) outcomes (92%-100% vs. 15%-23%). For at least eight of the RCTs, it was unclear whether or not the awareness of the THR type would influence the ascertainment of clinical/functional scores by patients/study personnel (performance bias) or outcome assessors (detection bias). Most RCTs failed to report blinding status of the patients, study personnel, and/or outcome assessors. For eight RCTs, the influence of attrition bias was judged at low risk. Five RCTs[112,121,122,124,125] were judged as being at high risk for selective outcome and/or analysis bias. Risk of other bias (e.g., funding source, balance imbalance in important characteristics, inappropriate analysis) for about one third of the RCTs was judged to be high.

Table 11. Risk of bias for RCTs: review author's judgments about each risk of bias item (THR vs. THR)

First author, year, study ID	**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
Angadi 2012[109]	+	+	?	+	?	+	+	+	+	?
Bjorgul 2010[107]	?	+	?	+	+	+	–	–	+	–
McCalden 2009[238]	?	?	?	+	+	+	+	+	+	–
Engh 2012[110]	?	?	?	?	?	+	–	–	+	?
Capello 2008[112]	?	?	?	+	?	+	+	+	–	–
Corten 2011[116]	+	+	+	+	+	+	+	+	+	+
Howie 2012[120]	+	+	NA	+	NA	+	NA	+	+	+
Lewis 2008[121]	?	+	?	+	?	+	?	?	–	?
Amanatullah 2011[122]	?	+	?	+	?	+	?	–	–	–
Kadar 2011[123]	?	+	+	+	?	+	+	+	+	+
Healy 2009[124]	–	–	?	+	–	+	+	+	–	+
Kim 2011[125]	+	?	?	+	?	+	+	+	–	+
Kim 2011[126]	?	+	?	+	?	+	+	+	+	?

ID=identification; ITT=intention-to-treat; PP=per protocol

Key:

– High risk of bias ? Unclear risk of bias + Low risk of bias NA Not applicable

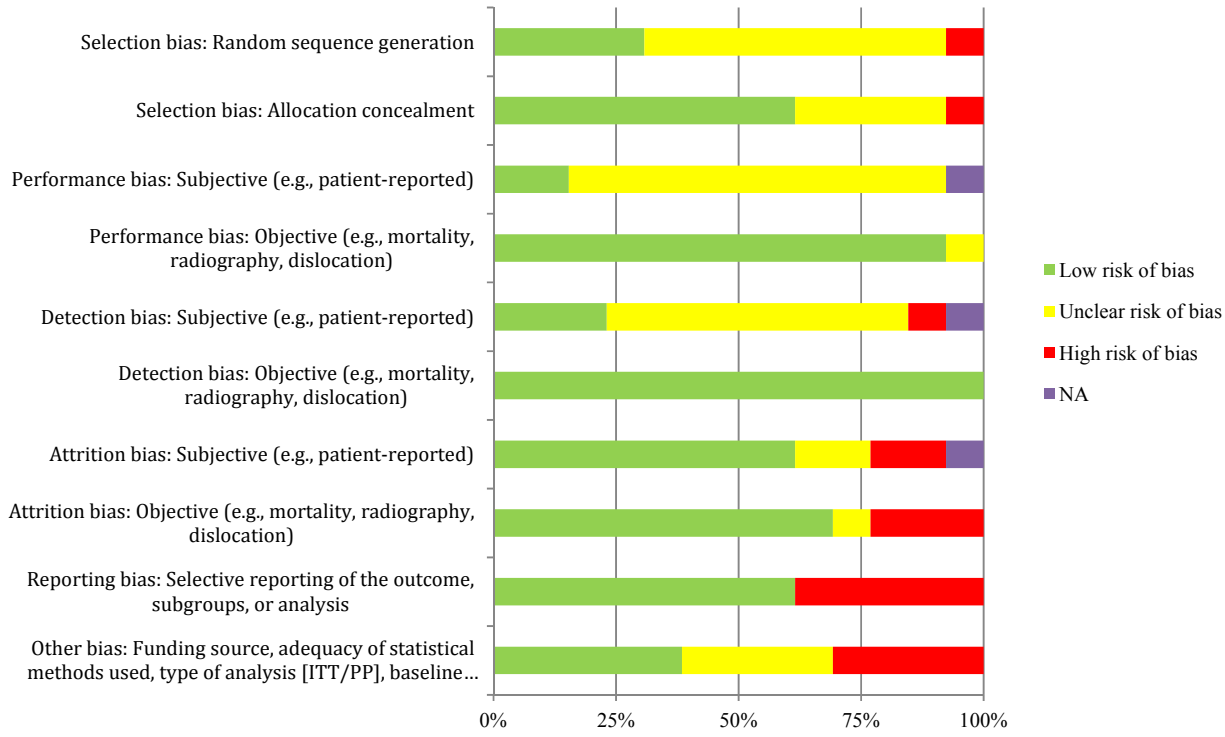


Figure 10. Risk of bias graph for RCTs: review author's judgments about each risk of bias item (THR vs. THR)

NA=not applicable; ITT=intention-to-treat; PP=per protocol

Methodological quality of systematic reviews

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 (Appendix 3). Briefly, based on the number of methodological items that were satisfied, two systematic reviews[134,137] were judged to be of high quality (falling into the score range of: 9-11) and two systematic reviews[135,138] were of medium quality (falling into the score range of: 5-8). The one remaining systematic review[136] had low quality (falling into the score range of: 0-4). The specific unmet methodological items were inappropriate analysis, absence of duplicate study selection, limited literature search, failure to address issues of publication bias, and no information on conflict of interest.

Table 12. Methodological quality assessment summary for systematic reviews (THR vs. THR)

First author, year, study ID	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**
Voigt 2012[134]	Yes	Yes	Yes	CA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High quality
Pakvis 2011[135]	Yes	No	Yes	Yes	No	Yes	Yes	No	No	No	No	Medium quality
Clement 2012[136]	Yes	No	No	Yes	Yes	Yes	No	No	No	No	No	Low quality
Sedrakyan 2011[137]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	High quality
Yoshitomi 2009[138]	Yes	Yes	Yes	CA	Yes	Yes	Yes	NA	No	Yes	No	Medium quality

ID=identification; NA= Not applicable; CA= Can't answer

6.2.2.3 Clinical effectiveness for THR vs. THR

This section summarises evidence on the findings from 13 RCTs[107,109,110,112,116,120-126,238] and five systematic reviews.[134-138]

The reported outcomes for this section were the following: Harris Hip score (12 RCTs[107,109,110,112,116,121-126,238]; 3 systematic reviews[134,135,137], WOMAC score (4 RCTs[116,121,126,238]), MACTAR score (1 RCT[116]), Merle d'Aubigne and Postel score (1 RCT[116]), UCLA score (1 RCT[126]), and Oxford Hip score (1 systematic review[134]), health related quality of life scale SF-12 (3 RCTs[121,122,238]; 1 systematic review[137]), risk of revision (10 RCTs[109,110,112,116,120-122,124-126]; 5 systematic reviews[134-138]), mortality (6 RCTs[107,110,116,120,125,238]), femoral head penetration rate (3 RCTs[110,123,238]), implant dislocation (7 RCTs[107,109,112,120-122,124]; 2 systematic reviews[136,137]), osteolysis (7 RCTs[109,110,112,122,124,126,238]; 2 systematic reviews[135,136]), aseptic loosening (5 RCTs[109,110,116,121,124]; 1 systematic review[136]), femoral fracture (3 RCTs[110,112,124]), infection (4 RCTs[109,121,122,124]), and deep vein thrombosis (1 RCT[122]).

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

- Hip Disability and Osteoarthritis Outcome Score (HOOS)
- Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH)
- American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire
- Arthritis Impact Measurement Scale (AIMS)
- Nottingham Health Profile questionnaire (NHP)
- Euro-Qol questionnaire (EQ-5D)
- Short Form 36 (SF-36) Health Survey
- Time to revision
- Pain score (visual analogue scale)

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

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 post-procedure follow-ups: Harris Hip score (12 studies),[107,109,110,112,116,121-126,238] WOMAC score (4 studies),[116,121,126,238] MACTAR score (1 study),[116] Merle d'Aubigne and Postel score (1 study),[116] and UCLA score (1 study).[126] None of these 12 studies reported measurements involving Oxford Hip score.

Three of the five included systematic reviews comparing different types of THR reported at least some evidence on Harris Hip score[134,135,137] and Oxford Hip score.[134] None of the three reviews reported any summary evidence for the WOMAC, MACTAR, Merle d'Aubigne and Postel, and UCLA scores.

Harris Hips score

RCTs (n=12)

Mean Harris scores at follow-up did not differ (range: 6 months to 10 years;

Table 13) between the following interventions: cup fixation (2 studies; cemented vs. cementless),[107,109] cup liner bearing surface (2 studies; cross-linked polyethylene vs. non cross-linked polyethylene),[110,238] cup and femoral stem fixation (1 study; cemented vs. cementless),[116] and femoral head-on-cup liner bearing surfaces (1 study; cobalt chromium/oxinium-on-polyethylene vs. cobalt chromium/oxinium-on-cross-linked polyethylene).[123] The pooled mean difference for Harris Hip score in our meta-analysis of two studies (Figure 11) comparing cup liners made with cross-linked polyethylene vs. non cross-linked polyethylene was 2.29 (95% CI: -0.88, 5.45),[110,238] suggesting a non-significant benefit of cross-linked polyethylene cup liners.

The evidence for the other comparisons based on cup shell design (porous coated vs. arc-deposited hydroxyapatite-coated),[112] femoral head bearing surface (oxinium vs. cobalt chromium),[121] femoral head-on-cup liner bearing surfaces (ceramic-on-ceramic vs. metal-on- polyethylene or ceramic-on-polyethylene),[112,122] femoral stem composition (cobalt chromium vs. titanium),[124] femoral stem design (short metaphyseal-fitting vs. conventional diaphyseal-filling),[125] and femoral stem fixation (cemented vs. cementless)[126] was also considered inconclusive by us.

Table 13. Harris Hip score (range: 0-100) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup fixation **Cemented vs. Cementless**				
6 mo 2 yrs 5 yrs 10 yrs 10 yrs	90.2 (87.9, 92.6) vs. 89.1 (86.9, 91.3)[107] 92.7 (89.6, 95.8) vs. 94.0 (92.4, 95.7)[107] 93.9 (91.6, 96.2) vs. 91.4 (89.3, 93.5)[107] 89.8 (87.0, 92.6) vs. 87.3 (84.1, 90.6)[107] 74.5 (NR) vs. 78.0 (NR)[109]	p>0.05 (NS) p>0.05 (NS) p>0.05 (NS) p>0.05 (NS) p>0.05 (NS)	2 [unclear ROB]	No difference
Cup liner bearing surface **XLPE vs. Non XLPE**				
1 yr 5 yrs 10 yrs	85.0 (10.3) vs. 83.4 (13.1)[238] 86.0 (13.1) vs. 83.1 (15.4)[238] 88.0 (14.0) vs. 86.0 (15.0)[110]	MD=1.60, 95% CI: -3.07, 6.27£ MD=2.90, 95% CI: -2.77, 8.57£ MD=2.00, 95% CI: -1.85, 5.85£ **Pooled estimate of MD**£ 2.29, 95% CI: -0.88, 5.45[110,238]	2 [unclear ROB]	No difference
Cup shell design **Porous-coated shell vs. Arc-deposited HA-coated shell**				
5 yrs 10 yrs	97.0 (NR) vs. 96.4 (NR)[112] 96.0 (NR) vs. 96.7 (NR)[112]	p>0.05 (NS) p>0.05 (NS)	1 [unclear ROB]	Inconclusive
Cup and femoral stem fixation **Cemented cup/femoral stem vs. Cementless cup/femoral stem**				
3 mo 6 mo 1 yr 3 yrs 5 yrs 7 yrs	41 (12.0) vs. 41 (11.0)[116] 47 (12) vs. 50 (13)[116] 52 (10.0) vs. 53 (11.0)[116] 50 (14.0) vs. 52 (11.0)[116] 47 (14.0) vs. 48 (13.0)[116] 44 (15) vs. 46 (14)[116]	MD=0.0, 95% CI: -3.00, 3.00£ MD=-3.0, 95% CI: -6.32, 0.32£ MD=-1.0, 95% CI: -3.86, 1.86£ MD=-2.0, 95% CI: -5.62, 1.62£ MD=-1.0, 95% CI: -4.88, 2.87£ MD=-2.0, 95% CI: -7.07, 3.05£	1 [low ROB]	No difference
Femoral head bearing surface **Oxinium femoral heads vs. CoCr femoral heads**				
2 yrs	92 (NR) vs. 92.5 (NR)[121]	p>0.159 (NS)	1 [unclear ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces-I **Ceramic-on-Ceramic vs. Metal-on-PE**				
5 yrs 10 yrs	96.4 (NR) vs. 97.0 (NR)[112] 96.7 (NR) vs. 96.4 (NR)[112]	p>0.05 (NS) p>0.05 (NS)	1 [unclear ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces -II **Ceramic-on-Ceramic vs. Ceramic-on-PE**				
5 yrs	NR[122]	p>0.05 (NS)	1 [unclear ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces–III **Steel-on-PE vs. CoCr-on-PE vs. Oxinium-on-PE vs. CoCr-on-XLPE vs. Oxinium-on-XLPE**				
2 yrs	91 (10.8) vs. 91 (8.5) vs. 91 (11.1) vs. 93 (11.3) vs. 88 (9.5)[123]	p=0.7 (NS) ANOVA-based p=0.5 (NS)£	1 [low ROB]	No difference

Femoral stem composition **CoCr vs. Titanium**				
5 yrs	83 (NR) vs. 87 (NR)[124]	p=0.029 (SS)	1 [high ROB]	Inconclusive
Femoral stem design **Short metaphyseal-fitting stem vs. Conventional metaphyseal- and diaphyseal-filling stem**				
3 yrs	97.0 (NR) vs. 96.0 (NR)[125]	p=0.79 (NS)	1 [unclear ROB]	Inconclusive
Femoral stem fixation **Cemented vs. Cementless**				
18 yrs	91 (NR) vs. 90 (NR)[126]	p=0.71(NS)	1 [unclear ROB]	Inconclusive

SROB=summary risk of bias; MD=mean difference; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene; PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome

Only those comparisons listed for which any evidence for the given outcome was reported
[One RCT – Howie 2012[120] did not report any evidence on this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based
£ Calculated

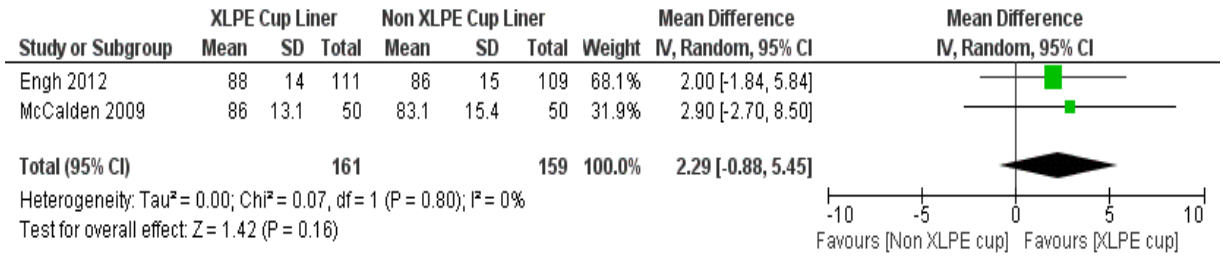


Figure 11. Harris Hip Score

Systematic reviews (n=3)

One systematic review reported the pooled mean difference for the HHS (Table 14).[137] Pooled estimates for the comparison of metal-on-metal vs. metal-on-polyethylene bearing surfaces for two different follow-ups were not consistent at two years; metal-on-metal gave significantly higher HHS than metal-on-polyethylene, but at over two years there was no significant difference between the two types of THR. The remaining two systematic reviews presented only narrative summaries.[134,135] In summary, for the HHS the systematic review-based evidence was considered inconclusive by us.

Table 14. Harris Hip score (range: 0-100) - Systematic reviews

Follow-up	Pooled effect estimate (95% CI)	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
3 yrs 2-5 yrs	NR [134] NR[135]	2[134] 3[135]	High quality [134] Low quality [135]	Inconclusive
Femoral head-on-cup liner surfaces-I Metal-on-Metal vs. Metal-on-PE				
2 yrs >2yrs	MD=-2.40, 95% CI: -4.47, -0.33 (SS)[137] MD=1.21, 95% CI: -2.41, 4.83 (NS)[137]	4[137] 2[137]	High quality [137]	Inconclusive
Femoral head-on-cup liner surfaces-II Ceramic-on-Ceramic vs. Ceramic-on-PE				
NR	NR[137]	5[137]	High quality [137]	Inconclusive
Femoral head-on-cup liner surfaces-III Ceramic-on-PE vs. Metal-on-PE				
NR	NR[137]	2[137]	High quality [137]	Inconclusive
Femoral head-on-cup liner surfaces-IV Metal-on-Metal vs. Ceramic-on-Ceramic				
NR	NR[137]	1[137]	High quality [137]	Inconclusive

MD=mean difference; 95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); PE=polyethylene; SS=statistically significant; NS=statistically not significant; MA=meta-analysis

Only those comparisons listed for which any evidence for the given outcome was reported
[Two systematic reviews Clement 2012[136] and Yoshitomi 2009[138] did not report this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive

Western Ontario and McMaster University Osteoarthritis Index (WOMAC) score

RCTs (n=4)

Results from all four RCTs reporting post-procedural 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),[238] cup and femoral stem fixation (cemented vs. cementless),[116] femoral head bearing surface (oxinium vs. cobalt chromium),[121] and femoral stem fixation (cemented vs. cementless) (Table 15).[126] The mean difference in WOMAC score of -0.12 (95%

CI: -7.58, 7.34) observed for one RCT[238] suggested no difference between cross-linked polyethylene and non-cross-linked polyethylene cup liners. Results on WOMAC score in the remaining three RCTs were considered inconclusive by us due to incompletely reported data.[116,121,126]

Table 15. The Western Ontario and McMaster University Osteoarthritis Index (range: 0-100) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup liner bearing surface XLPE vs. Non XLPE				
1 yr 5 yrs	83.0 (17.2) vs. 81.6 (17.6)[238] 78.0 (19.4) vs. 78.1 (18.2)[238]	MD=1.43, 95% CI: -5.48, 8.34[£] MD=-0.12, 95% CI: -7.58, 7.34[£]	1 [unclear ROB]	No difference
Cup and femoral stem fixation Cemented cup/femoral stem vs. Cementless cup/femoral stem				
NA	Mean domain sub-scores only[116]	-	1 [low ROB]	NA
Femoral head bearing surface Oxinium femoral heads vs. CoCr femoral heads				
2 yrs	84.9 (NR) vs. 87.0 (NR)[121]	p>0.159 (NS)	1 [unclear ROB]	Inconclusive
Femoral stem fixation Cemented vs. Cementless				
16 yrs	11 (NR) vs. 13 (NR)[126]	p=0.927(NS)	1 [unclear ROB]	Inconclusive

SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene; PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome; MD=mean difference

Only those comparisons listed for which any evidence for the given outcome was reported
[Nine RCTs – Bjorgul 2010[107,108] Angadi 2012[109] Engh 2012[110,111] Howie 2012[120] Capello 2008[112-115] Amanatullah 2011[122] Kadar 2011[123] Healy 2009[124] Kim 2011[125] did not report any evidence on this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based
[£] Calculated

Systematic reviews (n=0)

No evidence was identified.

Other functional/clinical scores

RCTs (n=2)

In one RCT, mean MACTAR scores (at 7 years: 0.20, 95% CI: -0.74, 1.14) and Merle d'Aubigne scores (at 7 years: -0.40, 95% CI: -1.34, 0.54), were not different in patients who received THR with either cemented or cementless components (Table 16, Table 17).[116] Results from one RCT comparing femoral stem fixation (cemented vs. cementless) on the post-operative UCLA score were inconclusive due to incomplete data reporting (Table 18).[126]

Table 16. The McMaster-Toronto Arthritis Patient Preference Disability Questionnaire score (range: 0-30) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup and femoral stem fixation Cemented cup/femoral stem vs. Cementless cup/femoral stem				
3 mo 6 mo 1 yr 3 yrs 5 yrs 7 yrs	**Mean change (post-operative)** -5.3 (2.5) vs. -5.2 (2.2)[116] -6.6 (1.9) vs. -6.4 (2.1)[116] -7.0 (1.8) vs. -6.9 (2.0)[116] -6.6 (2.3) vs. -6.4 (2.3)[116] -6.0 (2.8) vs. -6.2 (2.4)[116] -6.2 (2.8) vs. -6.0 (2.6)[116]	**Mean change difference** MD=0.10, 95% CI: -0.51, 0.71[£] MD=0.20, 95% CI: -0.33, 0.73[£] MD=0.10, 95% CI: -0.41, 0.61[£] MD=0.20, 95% CI: -0.46, 0.86[£] MD=-0.20, 95% CI: -0.45, 0.55[£] MD=0.20, 95% CI: -0.74, 1.14[£]	1 [low ROB]	No difference

SROB=summary risk of bias; MD=mean difference; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s)

Only those comparisons listed for which any evidence for the given outcome was reported
[None of the studies except for Corten 2011[116-119] reported any evidence on this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based
[£] Calculated

Table 17. Merle D'Aubigne and Postel score (range: 0-18) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup and femoral stem fixation Cemented cup/femoral stem vs. Cementless cup/femoral stem				
3 mo 6 mo 1 yr 3 yrs 5 yrs 7 yrs	**Mean change (post-operative)** 5.8 (1.9) vs. 5.6 (2.2)[116] 6.7 (2.1) vs. 7.0 (2.2)[116] 7.5 (1.8) vs. 7.4 (2.1)[116] 7.1 (2.2) vs. 6.9 (2.1)[116] 6.5 (2.3) vs. 6.6 (2.4)[116] 6.1 (2.6) vs. 6.5 (2.8)[116]	**Mean change difference** MD=0.20, 95% CI: -0.34, 0.74[£] MD=-0.30, 95% CI: -0.87, 0.27[£] MD=0.10, 95% CI: -0.43, 0.63[£] MD=0.20, 95% CI: -0.41, 0.81[£] MD=-0.10, 95% CI: -0.77, 0.57[£] MD=-0.40, 95% CI: -1.34, 0.54[£]	1 [low ROB]	No difference

SROB=summary risk of bias; MD=mean difference; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s)

Only those comparisons listed for which any evidence for the given outcome was reported
[None of the studies except for Corten 2011[116-119] reported any evidence on this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based
[£] Calculated

Table 18. The University of California, Los Angeles activity scale (range: 1-10) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Femoral stem fixation Cemented vs. Cementless				
16 yrs	7.6 (NR) vs. 7.8 (NR)[126]	p=0.814 (NS)	1 [unclear ROB]	Inconclusive

SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); MD=mean difference

Only those comparisons listed for which any evidence for the given outcome was reported
[None of the studies except for Kim 2011[126] reported any evidence on this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based

Systematic reviews (n=1)

OHS was reported in one systematic review comparing cup fixation methods (cemented vs. cementless), but was inconclusive (Table 19).[134] This evidence was based on one RCT showing a statistically non-significant result.

Table 19. Oxford Hip score (range: 0-48) - Systematic reviews

Follow-up	Pooled effect estimate (95% CI)	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
3 yrs	NR [134]	1[134]	High quality [134]	Inconclusive

MD=mean difference; 95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); PE=polyethylene; SS=statistically significant; NS=statistically not significant; MA=meta-analysis

Only those comparisons listed for which any evidence for the given outcome was reported
[None of the systematic reviews except for Voigt 2012[134] reported this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive

Health related quality of life

Only three RCTs[121,122,238] and one systematic review [137] reported any comparative evidence on measures of health related quality of life.

RCTs (n=3)

In one RCT, at follow-up of 1-5 years, there was no difference in quality of life (on mental and physical subscales of SF-12) between two THR groups of patients receiving cross-linked vs. non cross-linked polyethylene cup liner bearings (see Table **20**).[238]

In two other RCTs,[121,122] there were no statistically significant differences in the mean follow-up SF-12 mental and physical subscale scores between THR groups with different femoral head bearings (oxinium vs. cobalt chromium)[121] and femoral head-on-cup liner articulations (ceramic-on-ceramic vs. ceramic-on-polyethylene).[122] This evidence was considered to be inconclusive by us.

Table 20. Short Form Health Survey (SF-12; range: 0-100) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup liner bearing surface XLPE vs. Non XLPE				
1 yr 5 yrs	**Mental component score** 55.79 (7.38) vs. 56.01 (8.55)[238] **Physical component score** 42.20 (11.37) vs. 40.86 (11.11)[238] **Mental component score** 55.24 (8.01) vs. 53.36 (10.13)[238] **Physical component score** 37.24 (12.16) vs. 40.00 (11.78)[238]	**Mental component score** MD=-0.22, 95% CI: -3.38, 2.94[£] **Physical component score** MD=1.34, 95% CI: -3.12, 5.80[£] **Mental component score** MD=1.88, 95% CI: -1.74, 5.50[£] **Physical component score** MD=-2.76, 95% CI: -7.51, 1.99[£]	1 [unclear ROB]	No difference
Femoral head bearing surface Oxinium femoral heads vs. CoCr femoral heads				
2 yrs	**Mental component score** 53.80 (NR) vs. 52.57 (NR)[121] **Physical component score** 45.20 (NR) vs. 49.20 (NR)[121]	**Mental component score** p>0.05 (NS) **Physical component score** p>0.05 (NS)	1 [unclear ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces Ceramic-on-Ceramic vs. Ceramic-on-PE				
5 yrs	NR[122]	p>0.05 (NS)	1 [unclear ROB]	Inconclusive

SROB=summary risk of bias; MD=mean difference; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene; PE=polyethylene; CoCr= cobalt chrome

Only those comparisons listed for which any evidence for the given outcome was reported

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

** Decision was consensus-based

[£] Calculated

Systematic reviews (n=1)

One systematic review[137] reported two studies which 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 by us.

Table 21. Short Form Health Survey (SF-12; range: 0-100) - Systematic reviews

Follow-up	Pooled effect estimate (95% CI)	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Femoral head-on-cup liner surfaces-I Metal-on-Metal vs. Metal-on-PE				
2-3 yrs	NR[137]	2[137]	High quality [137]	Inconclusive

MD=mean difference; 95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); PE=polyethylene; SS=statistically significant; NS=statistically not significant; MA=meta-analysis

Only those comparisons listed for which any evidence for the given outcome was reported

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

Revision

Evidence on revision was reported for 10 randomised RCTs[109,110,112,116,120-122,124-126] and five systematic reviews.[134-138]

RCTs (n=10)

One RCT demonstrated a reduced risk of revision in patients who received cross-linked polyethylene vs. non cross-linked polyethylene cup liners (RR=0.18, 95% CI: 0.04, 0.78) (see Table 22).[110] The evidence reported in the remaining nine RCTs indicated statistically non-significant differences in risk of revision between the different types of THRs with wide confidence intervals compatible to large size effects in both directions (i.e., favouring one or the other treatment group). This evidence was also deemed inconclusive by us (see Table 22).

Table 22. Revision rate (n/N) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
10 yrs	17/183 vs. 11/104[109]	p>0.05 (NS); RR=0.87, 95% CI: 0.42, 1.80[£]	1 [low ROB]	Inconclusive
Cup liner bearing surface XLPE vs. Non XLPE				

10 yrs	2/111 vs. 11/109[110]	p<0.05 (SS); RR=0.18, 95% CI: 0.04, 0.78[£]	1 [unclear ROB]	In favour of XLPE cup liner
Cup shell design **Porous-coated shell vs. Arc-deposited HA-coated shell**				
5 yrs 5-10 yrs	2/113 vs. 4/109[112] 2/113 vs. 2/109[112]	p>0.05 (NS); RR=0.48, 95% CI: 0.09, 2.57[£] p>0.05 (NS); RR=0.96, 95% CI: 0.13, 6.72[£]	1 [low ROB]	Inconclusive
Cup and femoral stem fixation **Cemented cup/femoral stem vs. Cementless cup/femoral stem**				
7 yrs	13/124 vs. 6/126[116]	p=0.11 (NS); RR=2.20, 95% CI: 0.86, 5.60[£]	1 [low ROB]	Inconclusive#
Femoral head size **36 mm vs. 28 mm**				
1 yr	4/273 vs. 6/284[120]	p=NR; RR= 0.69, 95% CI: 0.19, 2.43[£]	1 [low ROB]	Inconclusive
Femoral head bearing surface **Oxinium femoral heads vs. CoCr femoral heads**				
2 yrs	1/50 vs. 1/50[121]	p=NR; RR= 1.00, 95% CI: 0.06, 15.50[£]	1 [low ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces-I **Ceramic-on-Ceramic vs. Metal-on-PE**				
5 yrs 5-10 yrs	6/222 vs. 8/106[112] 4/222 vs. 5/106[112]	p=0.045 (SS); RR= 0.35, 95% CI: 0.12, 1.00[£] p=0.08 (NS); RR= 0.38, 95% CI: 0.10, 1.39[£]	1 [low ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces -II **Ceramic-on-Ceramic vs. Ceramic-on-PE**				
5 yrs	11/196 vs. 3/161[122]	p=0.06 (NS); RR= 3.01, 95% CI: 0.85, 10.61[£]	1 [low ROB]	Inconclusive
Femoral stem composition **CoCr vs. Titanium**				
5 yrs	2/199 vs. 0/191[124]	p=0.16 (NS); RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive
Femoral stem design **Short metaphyseal-fitting stem vs. Conventional metaphyseal- and diaphyseal-filling stem**				
3 yrs	0/50 vs. 0/50[125]	p=NR; RR and 95% CI not estimated	1 [low ROB]	Inconclusive
Femoral stem fixation **Cemented vs. Cementless**				
20 yrs	Acetabular: 14/109 vs. 18/110[126] Femoral: 3/109 vs. 4/110[126]	p=0.673 (NS); RR= 0.78, 95% CI: 0.41, 1.49[£] p=0.912 (NS); RR= 0.75, 95% CI: 0.17, 3.30[£]	1 [low ROB]	Inconclusive

RR=risk ratio (relative risk); SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene; PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome

Only those comparisons listed for which any evidence for the given outcome was reported

[Three RCTs – Bjorgul 2010[107,108] McCalden 2009[238] and Kadar 2011[123] did not report any evidence on this outcome]

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

** Decision was consensus-based

[£] Calculated

The use of cementless implants (cup and femoral stem) was associated with better implant survival rate compared to cemented implants at 10 (83% vs. 94%, p=0.007), 15 (66% vs. 80%, p=0.007), and 20 years (48% vs. 69%, p=0.007) post-procedure[116] and was rated as in favour of cementless cup and femoral stem

Systematic reviews (n=5)

Of the five systematic reviews reporting on revisions, two provided pooled estimates for risk of revision (Table 23).[134,138] According to one review,[138] at 9 years post-surgery, the recipients of zirconium femoral head were at similar risk for revision compared to the recipients of non- zirconium femoral heads (3 pooled RCTs; RD=0.02, 95% CI: -0.01, 0.06). This evidence was considered conclusive in detecting no difference in revision rates between these two types of femoral head.

In another review,[134] 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, 1.18). This result was considered inconclusive by us given the uninformative 95% confidence intervals. Evidence from the remaining three reviews was of a narrative nature which precluded us drawing conclusions.[135-137]

Table 23. Revision rate (n/N) - Systematic reviews

Follow-up	Pooled effect estimate (95% CI)	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
4-8 yrs 10 yrs <10 yrs 5-15 yrs	RR=0.15, 95% CI: 0.02, 1.18 (NS)[134] RR=1.36, 95% CI: 0.81, 1.29 (NS)[134] NR[135] NR[136]	2[134] 2[134] 6[135] NR[136]	High quality [134] Low quality [135] Low quality [136]	Inconclusive
Femoral head-on-cup liner surfaces-I Metal-on-Metal vs. Metal-on-PE				
2-5 yrs	NR[137]	2[137]	High quality [137]	Inconclusive
Femoral head-on-cup liner surfaces-II Ceramic-on-Ceramic vs. Metal-on-PE				
6-8 yrs	NR[137]	1[137]	High quality [137]	Inconclusive
Femoral head-on-cup liner surfaces-III Ceramic-on-Ceramic vs. Ceramic-on-PE				
2-8 yrs	NR[137]	5[137]	High quality [137]	Inconclusive
Femoral head-on-cup liner surfaces-IV Ceramic-on-PE vs. Metal-on-PE				
8 yrs	NR[137]	1[137]	High quality [137]	Inconclusive
Femoral head-on-cup liner surfaces-V Zirconia-on-PE vs. Non Zirconia-on-PE				
9 yrs	RD=0.02, 95% CI: -0.01, 0.06 (NS)[138]	3[138]	Medium quality [138]	No difference

95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); PE=polyethylene; RD=risk difference; SS=statistically significant; NS=statistically not significant; MA=meta-analysis

Only those comparisons listed for which any evidence for the given outcome was reported

[All systematic reviews reported this outcome]

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

Mortality

The evidence on mortality was reported for six RCTs.[107,110,116,120,125,238] None of the five systematic reviews reported on mortality.

RCTs (n=6)

Evidence from the six RCTs that reported mortality was inconclusive due to non-significant RR estimates and wide 95% confidence intervals (see Table 24).[107,110,116,120,125,238] For example, based on a pooled RR estimate of 1.39 (95% CI: 0.78, 2.49),[110,238] 5-10 year post-surgery mortality rates between the recipients of cross-linked polyethylene vs. non cross-linked polyethylene cup liners were not significantly different (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),[107,116] and femoral head size (36 mm vs. 28 mm).[120] One RCT reported zero deaths for both treatment groups that received femoral stems of different design.[125]

Table 24. Mortality rate (n/N) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup fixation **Cemented vs. Cementless**				
10 yrs	12/107 vs. 14/108[107]	p=NR; RR=0.86, 95% CI: 0.41, 1.78[£]	1 [low ROB]	Inconclusive
Cup liner bearing surface **XLPE vs. Non XLPE**				
5 yrs 10 yrs	7/50 vs. 2/50[238] 17 /111 vs. 15/109[110]	p>0.05 (NS); RR=3.50, 95% CI: 0.76, 16.03[£] p>0.05 (NS); RR=1.11, 95% CI: 0.58, 2.11[£] **Pooled estimate of MH-RR** RR=1.39, 95% CI: 0.78, 2.49[110,238]	2 [unclear ROB]	Inconclusive
Cup and femoral stem fixation **Cemented cup/femoral stem vs. Cementless cup/femoral stem**				
7 yrs	18/124 vs. 17/126[116]	p=NR; 1.07, 95% CI: 0.58, 1.98[£]	1 [low ROB]	Inconclusive
Femoral head size **36 mm vs. 28 mm**				
1 yr	5/273 vs. 2/284[120]	p=NR; RR=2.58, 95% CI: 0.53, 13.20[£]	1 [low ROB]	Inconclusive
Femoral stem design **Short metaphyseal-fitting stem vs. Conventional metaphyseal- and diaphyseal-filling stem**				
3 yrs	0/50 vs. 0/50[125]	p=NR; RR and 95% CI not estimated	1 [low ROB]	Inconclusive

RR=risk ratio (relative risk); SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene; PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome

Only those comparisons listed for which any evidence for the given outcome was reported
[7 RCTs – Angadi 2012[109] Capello 2008[112-115] Lewis 2008[121] Amanatullah 2011[122] Kadar 2011[123] Healy 2009[124] and Kim 2011[126] did not report any evidence on this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive ** Decision was consensus-based £ Calculated

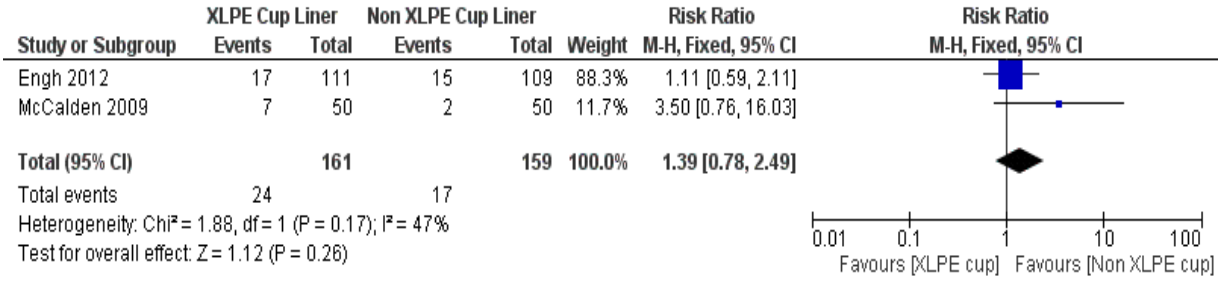


Figure 12. Mortality

Systematic reviews (n=0)

No evidence was identified.

Femoral head penetration rate (measure of prosthesis movement)

The evidence on femoral head penetration rate was reported for three RCTs.[110,123,238] None of the five systematic reviews reported this endpoint.

RCTs (n=3)

Two RCTs demonstrated reduced femoral head penetration in favour of cross-linked polyethylene vs. non cross-linked (conventional) polyethylene cup liners at 5-10 years of follow-up (

Table 25).[110,238] Similarly, in another RCT, cross-linked polyethylene cup liners with either metal or oxinium femoral heads outperformed conventional polyethylene cup liners in reducing femoral head penetration during two years of follow-up.[123]

Table 25. Femoral head penetration rate (mm/year) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup liner bearing surface XLPE vs. Non XLPE				
5 yrs 5 yrs 10 yrs	0.003 (-0.024, 0.030) vs. 0.051 (0.029, 0.073)[238] 0.24 (0.42) vs. 1.26 (0.62)[110] 0.06 (0.05) vs. 0.22 (0.11)[110]	p = 0.006 (SS) p<0.001 (SS) p<0.001 (SS)	2 [unclear ROB]	In favour of XLPE
Femoral head-on-cup liner bearing surfaces Steel-on-PE vs. CoCr-on-PE vs. Oxinium-on-PE vs. CoCr-on-XLPE vs. Oxinium-on-XLPE				
2 yrs	0.19 (0.16, 0.23) vs. 0.40 (0.33, 0.46) vs. 0.44 (0.37, 0.51) vs. 0.19 (0.15, 0.23) vs. 0.18 (0.13, 0.22)[123]	p<0.001 (SS; steel-PE, CoCr- XLPE, and Oxinium-XLPE vs. CoCr-PE and Oxinium-PE)	1 [low ROB]	In favour of CoCr-XLPE, Oxinium-XLPE, and steel-PE

SROB=summary risk of bias; MD=mean difference; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene; PE=polyethylene; CoCr= cobalt chrome

Only those comparisons listed for which any evidence for the given outcome was reported
[Three RCTs – McCalden 2009[238] Engh 2012[110] and Kadar 2011[123] reported this outcome]
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based

Systematic reviews (n=0)

No evidence was identified.

Complications

Evidence on the occurrence/absence of complications was reported by nine RCTs[109,110,112,120-122,124,126,238] and three systematic reviews.[135-137] In most studies,[109,110,112,120-122,126,238] reported complications were classified as post-operative. In one RCT,[124] some of the complications were classified as peri-operative.

Implant dislocation

RCTs (n=7)

Evidence on the occurrence/absence of implant dislocation was reported for seven RCTs (Table 26).[107,109,112,120-122,124] Our pooled estimate of two studies (Figure 13)[107,109] indicated a reduced risk of implant dislocation at 10 years follow-up in recipients of cemented vs. cementless cups (pooled OR=0.34, 95% CI: 0.13, 0.89). Moreover, in one RCT after one year of follow-up, the THR recipients with a larger size femoral head experienced a lower risk of implant dislocation compared to those with smaller size femoral head (36 mm vs. 28 mm; RR=0.17, 95% CI: 0.04, 0.78).[120] Evidence on implant dislocation for the remaining four RCTs[112,121,122,124] was inconclusive due to incomplete data and non-significant results.

Table 26. Implant dislocation rate (n/N) – RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup fixation **Cemented vs. Cementless**				
10 yrs	4/107 vs. 10/108[107] 1/183 vs. 3/104[109]	p>0.05 (NS); RR= 0.40, 95% CI: 0.13, 1.24[£] p=NR; RR= 0.18, 95% CI: 0.02, 1.79[£] **Pooled estimate of Peto OR**[£] OR=0.34, 95% CI: 0.13, 0.89[107,109]	2 [low ROB]	In favour of cemented cup
Cup shell design **Porous-coated shell vs. Arc-deposited HA-coated shell**				
10 yrs	2/113 vs. 3/109[112]	p=NR; RR= 0.64, 95% CI: 0.10, 3.77[£]	1 [low ROB]	Inconclusive
Femoral head size **36 mm vs. 28 mm**				
1 yr	2/258 vs. 12/275[120]	p=NR; RR= 0.17, 95% CI: 0.04, 0.78[£]	1 [low ROB]	In favour of 36 mm head size
Femoral head bearing surface **Oxinium femoral heads vs. CoCr femoral heads**				
2 yrs	2/50 vs. 1/50[121]	p=NR; RR= 2.00, 95% CI: 0.18, 21.35[£]	1 [low ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces -I **Ceramic-on-Ceramic vs. Ceramic-on-PE**				
5 yrs	10/166 vs. 9/146[122]	p=0.672 (NS); RR= 0.97, 95% CI: 0.40, 2.33[£]	1 [low ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces-II **Ceramic-on-Ceramic vs. Metal-on-PE**				
10 yrs	5/222 vs. 5/106[112]	p=0.25 (NS); RR=0.47, 95% CI: 0.14, 1.61[£]	1 [low ROB]	Inconclusive
Femoral stem composition **CoCr vs. Titanium**				
5 yrs	3/199 vs. 0/191[124]	p=0.678 (NS); RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive

RR=risk ratio (relative risk); OR=odds ratio; SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome

Only those comparisons listed for which any evidence for the given outcome was reported

[Six RCTs – McCalden 2009[238] Engh 2012[110] Corten 2011[116] Kadar 2011[123] Kim 2011[125] and Kim 2011[126] did not report any evidence on this outcome]

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

** Decision was consensus-based

[£] Calculated

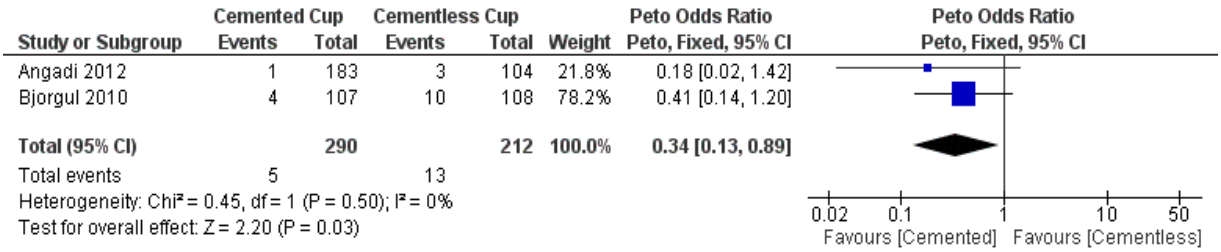


Figure 13. Implant dislocation

Systematic reviews (n=2)

Overall, no conclusions could be drawn from the two systematic reviews on implant dislocation, given the narrative evidence summary[137] and mixed study designs (Table 27).[136] The pooled data from one review[136] was based on nine studies most of which were not randomised and indicated a lower risk of dislocation in the groups receiving cemented vs. cementless cups.

Table 27. Implant dislocation rate (n/N) - Systematic reviews

Follow-up	Pooled effect estimate (95% CI)	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
5-15 yrs	12/914 (1.3%) vs. 28/696 (4.1%), p = 0.001[136] Pooled data from nine comparative studies (most non-RCTs) suggested that cemented cups had lower dislocation rate compared to cementless cups	NR[136]	Low quality[136]	Inconclusive
Femoral head-on-cup liner surfaces Metal-on-Metal vs. Metal-on-PE				
2-5 yrs	NR[137] No significant difference based on results from three RCTs	3[137]	High quality[137]	Inconclusive

95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); PE=polyethylene; MA=meta-analysis
Only those comparisons listed for which any evidence for the given outcome was reported
* Favours THR-1 (or THR-2), no difference, or inconclusive

Osteolysis

RCTs (n=7)

Evidence on osteolysis was reported by seven RCTs (Table 28).[109,110,112,122,124,126,238] In one RCT comparing different femoral head-on-cup liner bearing surfaces, recipients of ceramic-on-ceramic articulations had a reduced risk of osteolysis compared to recipients of metal-on-polyethylene articulations at 10 years post-operation (RR=0.10, 95% CI: 0.02, 0.32).[112]

For seven RCTs, the evidence for osteolysis was inconclusive across the comparisons based on different methods of cup fixation (cemented vs. cementless),[109] cup liner bearing surface (cross-linked polyethylene vs. non cross-linked polyethylene),[110,238] cup shell design (porous coated vs. arc-deposited hydroxylapatite-coated),[112] femoral head-on-cup liner bearing surface (ceramic-on-ceramic vs. ceramic-on-polyethylene),[122] femoral stem composition (cobalt chromium vs. titanium),[124] and femoral stem fixation (cemented vs. cementless).[126]

Table 28. Osteolysis (n/N) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
10 yrs	0/183 vs. 1/104[109]	p=NR; RR and 95% CI not estimated	1 [low ROB]	Inconclusive
Cup liner bearing service XLPE vs. Non XLPE				
5 yrs 10 yrs	0/50 vs. 0/50[238] 0/111 vs. 15/109[110]	p=NA; RR and 95% CI not estimated p<0.001; RR and 95% CI not estimated	2 [unclear ROB]	Inconclusive
Cup shell design Porous-coated shell vs. Arc-deposited HA-coated shell				
10 yrs	1/113 vs. 2/109[112]	p=NR; RR= 0.48, 95% CI: 0.04, 5.24[£]	1 [low ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces -I Ceramic-on-Ceramic vs. Ceramic-on-PE				
5 yrs	1/166 vs. 1/146[122]	p=0.797 (NS); RR= 0.87, 95% CI: 0.05, 13.93[£]	1 [low ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces-II Ceramic-on-Ceramic vs. Metal-on-PE				
10 yrs	3/222 vs. 15/106[112]	p<0.001 (SS); RR=0.10, 95% CI: 0.02, 0.32[£]	1 [low ROB]	In favour of Ceramic-on-Ceramic bearing surface
Femoral stem composition CoCr vs. Titanium				
5 yrs	0/199 vs. 0/191[124]	p=NR; RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive
Femoral stem fixation Cemented vs. Cementless				
20 yrs	Acetabular: 35/109 vs. 40/110[126] Femoral: 31/109 vs. 35/110[126]	p=0.168 (NS); RR=0.88, 95% CI: 0.61, 1.27[£] p=0.159 (NS); RR=0.89, 95% CI: 0.59, 1.33[£]	1 [low ROB]	Inconclusive

RR=risk ratio (relative risk); SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome; HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene

Only those comparisons listed for which any evidence for the given outcome was reported

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

** Decision was consensus-based

[£] Calculated

Systematic reviews (n=2)

Overall, no conclusions could be drawn on incidence of osteolysis from two low quality systematic reviews comparing cemented and cementless methods of cup fixation, given the narrative evidence summaries, mixed study designs, and inconsistent findings (Table 29).[135,136]

Table 29. Osteolysis (n/N) - Systematic reviews

Follow-up	Pooled effect estimate (95% CI)	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Cup fixation **Cemented vs. Cementless**				
2-6 yrs	NR[135] The analysis and narrative synthesis of RCT data showed no statistically significant difference in the occurrence of osteolysis between cemented and cementless cups.	3[135]	Low quality [135]	Inconclusive
5-15 yrs	NR[136] Narrative synthesis of nine comparative studies (most non-RCTs) indicated lower rates of osteolysis in cemented cups.	NR[136]	Low quality [136]	Inconclusive

95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); MA=meta-analysis
Only those comparisons listed for which any evidence for the given outcome was reported
* Favours THR-1 (or THR-2), no difference, or inconclusive

Other complications

RCTs (n=7)

Seven RCTs reported other complications such as aseptic loosening (see

Table **30**),[109,110,116,121,124] femoral fracture (see Table 31),[110,112,124] infection (see Table 32),[109,121,122,124] and deep vein thrombosis (see Table 33).[122] This evidence was judged to be inconclusive by us due to low event or zero event counts and confidence intervals indicating great uncertainty.

Table 30. Aseptic loosening (n/N) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
10 yrs	11/183 vs. 2/104[109]	p=NR; RR= 3.12, 95% CI: 0.70, 13.83[£]	1 [low ROB]	Inconclusive
Cup liner bearing surface XLPE vs. Non XLPE				
10 yrs	0/111 vs. 0/109[110]	NA; RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive
Cup and femoral stem fixation Cemented cup/femoral stem vs. Cementless cup/femoral stem				
20 yrs	9/124 vs. 4/126[116]	p=NR; RR= 2.28, 95% CI: 0.72, 7.23[£]	1 [low ROB]	Inconclusive
Femoral head bearing surface Oxinium femoral heads vs. CoCr femoral heads				
2 yrs	0/50 vs. 1/50[121]	p=NR; RR and 95% CI not estimated	1 [low ROB]	Inconclusive
Femoral stem composition CoCr vs. Titanium				
5 yrs	1/199 vs. 0/191[124]	p=0.324 (NS); RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive

RR=risk ratio (relative risk); SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome; HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene

Only those comparisons listed for which any evidence for the given outcome was reported
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based
[£] Calculated

Table 31. Femoral fracture (n/N) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup liner bearing service XLPE vs. Non XLPE				
10 yrs	2/111 vs. 0/109[110]	p=NR; RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive
Cup shell design Porous-coated shell vs. Arc-deposited HA-coated shell				
10 yrs	0/113 vs. 0/109[112]	NA; RR and 95% CI not estimated	1 [low ROB]	Inconclusive
Femoral stem composition CoCr vs. Titanium				
5 yrs	0/199 vs. 1/191[124]	p=0.309 (NS); RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive

RR=risk ratio (relative risk); SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome; HXLPE=highly cross-linked polyethylene; XLPE= cross-linked polyethylene

Only those comparisons listed for which any evidence for the given outcome was reported
* Favours THR-1 (or THR-2), no difference, or inconclusive
** Decision was consensus-based

Table 32. Infection (n/N) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
10 yrs	0/183 vs. 2/104[109]	p=NR; RR and 95% CI not estimated	1 [low ROB]	Inconclusive
Femoral head bearing surface Oxinium femoral heads vs. CoCr femoral heads				
2 yrs	1/50 vs. 1/50[121]	p=NR; RR= 1.00, 95% CI: 0.06, 15.55[£]	1 [low ROB]	Inconclusive
Femoral head-on-cup liner bearing surfaces Ceramic-on-Ceramic vs. Ceramic-on-PE				
5 yrs	Superficial: 6/166 vs. 3/146[122] Deep: 1/166 vs. 2/146[122]	p=0.357 (NS); RR= 1.75, 95% CI: 0.44, 6.90[£] p=0.909 (NS); RR= 0.43, 95% CI: 0.04, 4.79[£]	1 [low ROB]	Inconclusive
Femoral stem composition CoCr vs. Titanium				
5 yrs	1/199 vs. 0/191[124]	p=0.324 (NS); RR and 95% CI not estimated	1 [unclear ROB]	Inconclusive

RR=risk ratio (relative risk); SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); PE=polyethylene; CoCr= cobalt chrome

Only those comparisons listed for which any evidence for the given outcome was reported

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

** Decision was consensus-based

[£] Calculated

Table 33. Deep vein thrombosis (n/N) - RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI)	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Femoral head-on-cup liner bearing surfaces Ceramic-on-Ceramic vs. Ceramic-on-PE				
5 yrs	3/166 vs. 2/146[122]	p=0.909 (NS); RR= 1.31, 95% CI: 0.22, 7.78[£]	1 [low ROB]	Inconclusive

RR=risk ratio (relative risk); SROB=summary risk of bias; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s); PE=polyethylene

Only those comparisons listed for which any evidence for the given outcome was reported

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

** Decision was consensus-based

[£] Calculated

Systematic reviews (n=1)

Of other complications, only aseptic loosening was reported in one low quality systematic review (Table 34).[136] Pooled data from 11 studies, most of which were not randomised, pointed towards a greater risk of aseptic loosening with cemented vs. cementless cup, however, the evidence is inconclusive given the lack of numerical data and evidence synthesis based on mixed study designs.

Table 34. Aseptic loosening (n/N) - Systematic reviews

Follow-up	Pooled effect estimate (95% CI)	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Cup fixation Cemented vs. Cementless				
5-15 yrs	NR[136] Pooled data from 11 comparative studies (most non-RCTs) presented only graphically suggested higher rates of aseptic loosening with cemented vs. cementless cup.	NR[136]	Low quality [136]	Inconclusive

95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); MA=meta-analysis
Only those comparisons listed for which any evidence for the given outcome was reported
* Favours THR-1 (or THR-2), no difference, or inconclusive

6.2.3 Grading overall quality of evidence

The results for graded outcomes are presented in the following Evidence Profile (EP) (see Table 35). For a meaningful grading process and for consistency, only the THR comparison categories which included at least two studies (cup fixation – cemented vs. cementless and cup liner bearing surface: XLPE vs. non XLPE) 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, respectively), revision (very low grade), mortality (very low and low grade, respectively), femoral head penetration (not graded and moderate, respectively), and implant dislocation (high and not graded, respectively).

Table 35. GRADE evidence profile for gradable outcomes reported in RCTs of THR
(adapted from Guyatt et al., 2011)[99]

Outcome [follow-up timing]	N 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)*
Cup fixation (cemented vs. cementless) – 2 RCTs[107,109]								
Harris Hip score [6 mo-10 yrs]	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 yrs]	1 (287)	None **Inconclusive**	Low	NA	Direct	Imprecise	Likely	Very low
Mortality [10 yrs]	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 yrs]	2 (502)	OR=0.34 95% CI: 0.13, 0.89 **In favour** of **cemented cup**	Low	Consistent	Direct	Precise	Unlikely	High
Cup liner bearing surface (XLPE vs. Non XLPE) – 2 RCTs[110,238]								
Harris Hip score [1-10 yrs]	2 (320)	MD=2.29 95% CI:-0.88, 5.45 **No difference**	Unclear	Consistent	Direct	Precise	Unlikely	Moderate
WOMAC score [1-5 yrs]	1 (100)	None **No difference**	Unclear	NA	Direct	Precise	Likely	Very low
Revision [10 yrs]	1 (220)	None **In favour of XLPE cup liner**	Unclear	NA	Direct	Precise	Likely	Very low
Mortality [5-10 yrs]	2 (320)	RR=1.39 95% CI: 0.78, 2.49 **Inconclusive**	Unclear	Consistent	Direct	Imprecise	Unlikely	Low
Femoral head penetration [5-10 yrs]	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)

GRADE= Grading of Recommendations, Assessment, Development, and Evaluation; RCT=randomised controlled trial; CI=confidence interval; SROB=summary risk of bias; RCT=randomised controlled trial; NA=not applicable; yr(s)=year(s); mo(s)=month(s); THR=total hip replacement

*GRADE categories: high, moderate, low, very low, NA (no evidence)

6.2.4 Summary conclusions comparing THRs

RCTs

The majority of the evidence comparing THRs was rated as inconclusive by us (

Table **36**). In three RCTs comparing the use of a cemented cup (vs. cementless cup)[107,109] or larger femoral head size (36 mm vs. 28 mm),[120] there was evidence of a reduced risk of implant dislocation (high grade evidence for the cup fixation comparison). In three other RCTs, patients who received THR with XLPE cup liners experienced reduced (i.e., improved) femoral head penetration rate (moderate grade evidence)[110,123,238] and risk for revision (very low grade evidence)[110] compared to recipients of conventional PE cup liners. In one RCT, the use of cementless fixation of cup and femoral stem (vs. cemented fixation) were associated with better implant survival rate.[116] Moreover, the recipients of ceramic-on-ceramic articulations (vs. metal-on-polyethylene) experienced a reduced risk of osteolysis.[112] For half of the studies, the mean post-THR clinical and functional scores (i.e., HHS, WOMAC score, SF-12, MACTAR score, Merle D'Aubigne Postel score) measured at different follow-ups 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).[107,109,110,116,123,238]

Evidence from studies reporting the UCLA score,[126] mortality (very low grade evidence),[107,110,116,120,125,238] aseptic loosening,[109,110,116,121,124] femoral fracture,[110,112,124] infection,[109,121,122,124] and deep vein thrombosis[122] was all inconclusive. Also, the evidence reported in four studies was considered inconclusive for all outcomes (very low grade evidence).[121,122,124,125] Results were considered inconclusive by us due to the partial reporting (missing data for effect estimates, confidence intervals, standard errors, standard deviations, p-values), great uncertainty (wide confidence intervals), zero event counts, and/or inconsistency in estimates.

Table 36. Summary of evidence regarding the differences between the different types of THR for each reported outcome (RCTs)

Conclusive evidence suggesting difference	Conclusive evidence suggesting no difference	Inconclusive evidence
Cup fixation Cemented vs. Cementless[107,109]		
Implant dislocation [high grade evidence][107,109] **In favour of cemented**	Harris Hip score [moderate grade evidence][107,109]	Mortality [very low grade evidence][107] Revision [very low grade evidence][109] Osteolysis[109] Aseptic loosening[109] Infection[109]
Cup liner bearing surface XLPE vs. Non XLPE[110,238]		
Femoral head penetration [moderate grade evidence][110,238] Revision rate [very low grade evidence][110] **In favour of XLPE**	Harris Hip score [moderate grade evidence][110,238] WOMAC score [very low grade evidence][238] SF-12 (mental/physical)[238]	Mortality [low grade evidence][110,238] Aseptic loosening[110] Femoral fracture[110]
Cup shell design Porous-coated vs. Arc-deposited HA-coated[112]		
None	None	Harris Hip score Revision Implant dislocation Osteolysis Femoral fracture
Cup and femoral stem fixation Cemented vs. Cementless[116]		
None#	Harris Hip score Merle D'Aubigne Postel score MACTAR score	WOMAC score Mortality Revision Aseptic loosening
Femoral head size 36 mm vs. 28 mm[120]		
Implant dislocation **In favour of 36 mm**	None	Mortality Revision
Femoral head bearing surface Oxinium vs. CoCr[121]		
None	None	Harris Hip score SF-12 WOMAC score Revision Implant dislocation Aseptic loosening Infection
Femoral head-on-cup liner bearing-I Ceramic-on-Ceramic vs. Metal-on-PE[112]		
Osteolysis **In favour of ceramic-on-ceramic**	None	Harris Hip score Revision Implant dislocation

Femoral head-on-cup liner bearing-II Ceramic-on-Ceramic vs. Ceramic-on-PE[122]		
None	None	Harris Hip score SF-12 Revision Implant dislocation Osteolysis Infection Deep vein thrombosis
Femoral head-on-cup liner bearing-III Steel-on-PE vs. CoCr/Oxinium-on-XLPE vs. CoCr/Oxinium-on-PE[123]		
Femoral head penetration **In favour of Steel-on-PE or CoCr/Oxinium-on-XLPE**	Harris Hip score	None
Femoral stem composition CoCr vs. Titanium[124]		
None	None	Harris Hip score Revision Implant dislocation Osteolysis Aseptic loosening Femoral fracture Infection
Femoral stem design Short metaphyseal-fitting vs. Conventional metaphyseal- and diaphyseal-filling[125]		
None	None	Harris Hip score Mortality Revision
Femoral stem fixation Cemented vs. Cementless[126]		
None	None	Harris Hip score UCLA score WOMAC score Revision Osteolysis

XLPE= cross-linked polyethylene; PE=polyethylene; HA=hydroxylapatite; CoCr= cobalt chrome; WOMAC=Western Ontario and McMaster University Osteoarthritis Index; SF-12=Short Form Health Survey; RCT=randomised controlled trial; UCLA= University of California, Los Angeles activity scale
# Implant survival rate was in favour of cementless[116]

Systematic reviews

Most evidence from the five systematic reviews comparing different types of THR[134-138] was considered inconclusive by us due to unreported pooled results across RCTs (i.e., reporting only narrative syntheses), reporting inappropriate pooling methods (e.g., indirect naïve comparison of single group cohorts; pooling of studies of different design),[135,136,138] or reporting of inconsistent summary findings (Table 37).[137] The evidence from one review indicated no difference in the risk for revision between two different articulations of zirconinum-on-polyethylene vs. non zirconium-on-polyethylene.[138]

Table 37. Summary of evidence regarding the differences between the compared types of THR for each reported outcome (systematic reviews)

Conclusive evidence suggesting difference	Conclusive evidence suggesting no difference	Inconclusive evidence
Cup fixation Cemented vs. Cementless[134-136]		
None	None	Harris Hip score[134,135] Oxford Hip score[134] Revision[134-136] Aseptic loosening[136]
Femoral head-on-cup liner bearing Different comparisons*[137,138]		
None	Revision[138]	Harris Hip score[137] SF-12[137] Revision[137] Implant dislocation[137]
PE=polyethylene		

*Metal-on-Metal vs. Metal-on-PE[137]
Ceramic-on-Ceramic vs. Ceramic-on-PE[137]
Ceramic-on-PE vs. Metal-on-PE[137]
Metal-on-Metal vs. Ceramic-on-Ceramic[137]
Zirconia-on-PE vs. Non Zirconia-on-PE[138]

6.2.4.1 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 due to an insufficient number of data points in the forest/funnel plots.

Heterogeneity, subgroup effects, and sensitivity analysis

The data reviewed from RCTs was 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 (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.

6.2.5 Comparison of total hip replacement and resurfacing arthroplasty (THR vs. RS)

6.2.5.1 Study and participant characteristics

RCTs

Study and participant characteristics of the three included RCTs[127-129] are summarised in Table 38. More details can be found in Appendices 4 & 5. Two RCTs were conducted in Canada[128,129] and one in the

UK.[127] A total of 422 participants were randomised across the three RCTs ranging from 104[128] to 192 participants.[129] Mean age ranged from 50[129] to 56 years,[127] and the proportion of women across the studies ranged from 10.5%[128] to 41%.[127] Length of follow of the studies ranged from one year[127] to six years.[129] The proportion of participants diagnosed with primary osteoarthritis was reported for two studies[127,129] and ranged from 33%[129] to 95%.[127]

Table 38. Overall study characteristic across three RCTs comparing THR vs. RS

Study Characteristic	Frequency
Geographical region	UK (n = 1); Canada (n = 2)
Total number of randomised participants	422 (range: 104 - 192)
Mean age (in years)	Range: 50 - 56
Female participants (%)	Range: 10.5 - 41
Length of follow-up (in years)	Range: 1 - 6
Diagnosis of primary osteoarthritis (%)	Range: 33 - 95

The three RCTs reported on clinical/functional scores (e.g., Harris Hip, Oxford Hip, UCLA, WOMAC scores), health related quality of life, and risk of revision. Follow-up of outcome assessments ranged from three weeks[127] to five years.[129] Outcomes reported in the included studies can be found in Appendix 9.

Systematic reviews

Three systematic reviews[139-141] were included which evaluated clinical effectiveness of THR compared to RS with respect to post-operative clinical/function (Harris Hip score, WOMAC score), risk of revision, mortality, and complications.[139,140] Searches for these systematic reviews were undertaken between March 2008[141] and January 2010.[140] Evidence was synthesized from both RCTs and non-RCTs (see Appendices 4 & 5). Further details on specific outcomes reported (or not reported) in the included systematic reviews can be found in Appendix 9.

6.2.5.2 Risk of bias and methodological quality

Risk of bias in RCTs

Risk of bias assessment for the three included RCTs[127-129] comparing THR to RS is presented in risk of bias tables (Appendix 3), the summary table (Table 39) and the risk of bias graph (Figure 14). Overall, two studies[127,129] reported an adequate method for random sequence generation and all three studies[127-129] reported treatment allocation concealment (low risk of bias). Two of the three studies[127,129] were rated as having 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,[128] but in the two other studies blinding of patients and study personnel was not undertaken.[127,129] For two studies, the influence of attrition bias on objective outcomes was judged at low risk.[127,129] All three studies were judged as being at low risk for selective outcome and/or analysis bias. Risk of other bias (e.g., funding source, balance/imbalance in important characteristics, inappropriate analysis) for one of the three studies was judged to be high.[128]

Table 39. Risk of bias summary for RCTs: review author's judgements about each risk of bias item (THR vs. RS)

First author, year, study ID	**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[127]	+	+	−	+	+	+	+	+	+	+
Garbuz 2010[128]	?	+	+	NA	?	NA	+	NA	+	−
Vendittoli 2010[129]	+	+	−	+	?	+	−	+	+	+

ID=identification; ITT=intention-to-treat; PP=per protocol

Key:

− High risk of bias ? Unclear risk of bias + Low risk of bias NA Not applicable

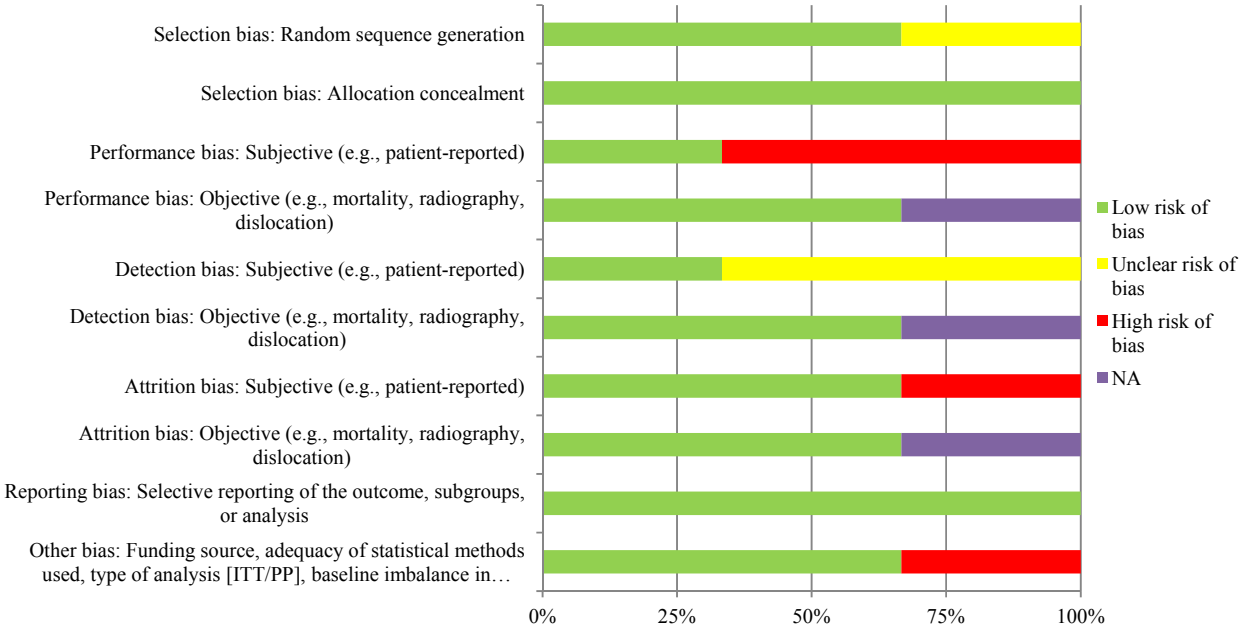


Figure 14. Risk of bias graph for RCTs: review author's judgements about each risk of bias item (THR vs. RS)
NA=not applicable; ITT=intention-to-treat; PP=per protocol

Methodological quality of systematic reviews comparing THR to RS

The assessment of methodological quality of the three included systematic reviews,[139-141] is presented in

Table **40** and the data extraction sheets (Appendices 4 & 5). Given the number of methodological items that were satisfied, the quality of one of the three reviews was judged as high (falling into the score range of: 9-11),[140] one as medium (falling into the score range of: 5-8),[139] and one as low (falling into the score range: 0-4).[141] The specific unmet methodological items were inappropriate analysis, failure to address issues of publication bias, and no information on conflict of interest.

Table 40. Methodological quality assessment summary for systematic reviews (THR vs. RS)

First author, year, study ID	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[139]	Yes	Yes	Yes	No	Yes	No	Yes	CA	No	No	No	Medium quality
Smith 2010[140]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CA	No	Yes	Yes	High quality
Springer 2009[141]	Yes	Yes	No	No	Yes	Yes	CA	No	No	No	No	Low quality

ID=identification; NA= Not applicable; CA= Can't answer

6.2.5.3 Clinical effectiveness for THR vs. RS

This section summarises evidence on the findings from three RCTs[127-129] and three systematic reviews. [139-141]

The reported outcomes for this section were the following: HHS (1 RCT;[127] 2 systematic reviews[139,140]), WOMAC score (2 RCTs;[128,129] 2 systematic reviews[139,140]), Merle d'Aubigne and Postel score (1 RCT;[129] 1 systematic review[139]), UCLA score (2 RCTs;[128,129] 1 systematic review[139]), OHS (1 RCT[127]), health related quality of life scales (SF-36 and Euro-Qol EQ-5D; 2 RCTs[127,128]), risk of revision (1 RCT[129]; 2 systematic reviews[139,140]), mortality (2 systematic reviews[139,140]), infection (2 RCTs;[127,129] 1 systematic review[139]), aseptic loosening (1 RCT;[129] 2 systematic reviews[139,140]), implant dislocation (2 RCTs;[127,129] 1 systematic review[139]), and deep vein thrombosis (2 RCTs[127,129]).

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

- Hip Disability and Osteoarthritis Outcome Score (HOOS)
- Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH)
- American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire
- Arthritis Impact Measurement Scale (AIMS)
- McMaster-Toronto Arthritis patient Preference Disability Questionnaire (MACTAR)
- Nottingham Health Profile questionnaire (NHP)
- Short Form (SF-12) Health Survey

- Time to revision
- Pain score (visual analogue scale)
- Femoral head penetration

Summary results for the outcomes listed below are presented separately for RCTs (

Table **41**) and systematic reviews (Table 42).

Evidence from RCTs

Functional/clinical measures

All three included RCTs comparing THR to RS reported some evidence for the following functional scores measured at 12-24 months after the procedure: HHS,[127] OHS,[127] WOMAC score,[128,129] UCLA score,[128,129] and Merle d'Aubigne and Postel score.[129]

In two RCTs, there was no difference between THR and RS groups in the mean post-operative OHS (12 months; MD=-2.23, 95% CI: -5.98, 1.52)[127] Merle d'Aubigne and Postel (24 months; MD=0.0, 95% CI: -1.06, 1.06),[129] and WOMAC scores (12 months; MD=2.20, 95% CI: -1.57, 5.97).[129] And although one of these RCTs at 24 months of follow-up showed a significantly improved mean WOMAC score for the RS group compared to the THR group, this difference was not deemed as clinically important (MD=3.30, 95% CI: 0.01, 6.58).[129]

There was inconclusive evidence for the three RCTs regarding the post-operative between-treatment differences with regards to the mean HHS and UCLA score[127] and incomplete data reporting,[128,129] respectively.

Health related quality of life

Two RCTs reporting quality of life measures showed statistically non-significant differences between THR and RS groups of patients for both SF-36 (p=0.55 and p=0.97 on mental and physical components, respectively)[128] and Euro-Qol EQ-5D (MD=-0.08, 95% CI: -0.18, 0.03).[127] These results were considered as inconclusive given the wide confidence interval[127] and incomplete data reporting.[128]

Revision

The occurrence of implant revision was reported for only one RCT.[129] There was no statistically significant difference between the THR and RS groups of patients in risk of revision at six (RR=1.01, 95% CI: 0.06, 15.92), 24 (RR=0.50, 95% CI: 0.04, 5.48), and 56 (RR=0.54, 95% CI: 0.10, 2.91) months post-surgery. The 95% confidence intervals 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 to RS.

Mortality rate

No evidence was identified from RCTs.

Complications

Evidence on complications was reported for two RCTs.[127,129] One meta-analysis of two RCTs (

Figure 15) indicated that at 12-56 months post-operation, THR recipients were at increased risk of infection compared to RS recipients (pooled OR=7.94, 95% CI: 1.78, 35.40).[127,129]

Evidence for the differences in the risk of deep vein thrombosis (

Figure **16**; pooled OR=0.60, 95% CI: 0.15, 2.42),[127,129] implant dislocation (

Figure 17 ; pooled OR=3.97, 95% CI: 0.79, 19.90),[127,129] wound complication (RR=4.01, 95% CI: 0.92, 18.18),[127] and aseptic loosening (RR not estimable)[129] was judged to be inconclusive by us.

Table 41. Summary results for total hip replacement vs. resurfacing arthroplasty – RCTs

Follow-up	Arm-specific estimates n/N or mean (SD or 95% CI) [THR vs. RS]	Difference (p value or 95% CI)	# of RCTs [SROB across studies]**	Treatment effect Conclusion*
Harris Hip score (range: 0-100)				
12 mo	82.3 (77.2, 87.5) vs. 88.4 (84.4, 92.4)[127]	MD=-6.04, 95% CI:- 12.58, 0.51	1 [low ROB]	Inconclusive
Oxford Hip score (range: 0-48)				
12 mo	38.2 (35.3, 41.0) vs. 40.4 (37.9, 42.9)[127]	MD=-2.23, 95% CI: (-5.98, 1.52)	1 [low ROB]	No difference
Western Ontario and McMaster University Osteoarthritis Index score (range: 0-100)				
3 mo 6 mo 12 mo 12 mo 24 mo	19.2 (NR) vs. 19.9 (NR)[129] 11.3 (NR) vs. 13.9 (NR)[129] 10.2 (10.7) vs. 8.0 (13.2)[129] 90.18 (NR) vs. 90.40 (NR)[128] 9.0 (11.9) vs. 5.7 (8.6)[129]	p=0.76 (NS)[129,133] p=0.20 (NS)[129,133] MD=2.20, 95% CI: -1.57, 5.97[129,133£] p=0.95 (NS)[128] MD=3.30, 95% CI: 0.01, 6.58[129,133£]	2 [unclear ROB]	No difference
Merle d'Aubigne and Postel score (range: 0-18)				
3 mo 6 mo 12 mo 24 mo	15.8 (NR) vs. 16.2 (NR)[129] 17.1 (NR) vs. 17.2 (NR)[129] 16.6 (NR) vs. 16.7 (NR)[129] 17.5 (1.3) vs. 17.5 (1.3)[129]	p=0.59 (NS) p=0.72 (NS) p=0.94 (NS) p=0.94 (NS); MD=0.0, 95% CI: -1.06, 1.06[£]	1 [unclear ROB]	No difference
University of California, Los Angeles activity score (range: 1-10)				
12 mo 12 mo 24 mo	6.3 (NR) vs. 6.8 (NR)[128] 6.3 (NR) vs. 7.1 (NR)[129] NR (NR) vs. NR (NR)[129]	p=0.24 (NS)[128] p=0.03 (SS)[129,133] p=0.09 (NS)[129,133]	2 [unclear ROB]	Inconclusive
Short Form-36 Health Survey (range: 0-100)				
12 mo 12 mo	**Mental component** 55.13 (NR) vs. 53.87 (NR)[128] **Physical component** 51.28 (NR) vs. 51.22 (NR)[128]	**Mental component** p= 0.55 (NS) **Physical component** p=0.97 (NS)	1 [unclear ROB]	Inconclusive
Euro-Qol [EQ-5D] questionnaire (range: 0-1)				
12 mo	0.71 (0.63, 0.80) vs. 0.79 (0.72, 0.87)[127]	MD=-0.077, 95% CI:-0.188, 0.034	1 [low ROB]	Inconclusive
Revision rate (n/N)				
3 mo 6 mo 12 mo 24 mo 56 mo	1/102 vs. 0/103[129] 1/102 vs. 1/103[129] 1/102 vs. 2/103[129] 1/102 vs. 2/103[129]	p=NR; RR and 95% CI not estimated p=NR; RR=1.01, 95% CI: 0.06, 15.92[£] p=NR; RR=0.50, 95% CI: 0.04, 5.48[£] p=NR; RR=0.50, 95% CI: 0.04, 5.48[£]	1 [low ROB]	Inconclusive

	2/100 vs. 4/109[129]	p=0.47 (NS); RR=0.54, 95% CI: 0.10, 2.91[£]		
Complications (n/N)				
Infection				
12 mo 56 mo	2/66 vs. 0/60[127] 5/100 vs. 0/109[129]	p=0.49 (NS); RR and 95% CI not estimated p=0.02 (SS); RR and 95% CI not estimated **Pooled estimate of Peto OR**[£] OR=7.94, 95% CI: 1.78, 35.40[127,129]	2 [low ROB]	In favour of RS
Deep vein thrombosis				
12 mo 56 mo	0/66 vs. 4/60[127] 3/100 vs. 1/109[129]	p=0.05 (NS); RR and 95% CI not estimated p=NR (NS); RR=3.27, 95% CI: 0.30, 30.90[£] **Pooled estimate of Peto OR**[£] OR=0.60, 95% CI: 0.15, 2.42[127,129]	2 [low ROB]	Inconclusive
Implant dislocation				
12 mo 56 mo	1/66 vs. 1/60[127] 4/100 vs. 0/109[129]	p=1.00 (NS); RR=0.90, 95% CI: 0.05, 14.21[£] p=0.038 (SS); RR and 95% CI not estimated **Pooled estimate of Peto OR**[£] OR=3.97, 95% CI: 0.79, 19.90[127,129]	2 [low ROB]	Inconclusive
Superficial wound complication				
12 mo	9/66 vs. 2/60[127]	p=0.06 (NS); RR=4.01, 95% CI: 0.92, 18.18[£]	1 [low ROB]	Inconclusive
Aseptic loosening				
56 mo	0/100 vs. 6/109[129]	p=0.017 (SS); RR and 95% CI not estimated	1 [low ROB]	Inconclusive

THR=total hip replacement; RS=resurfacing arthroplasty; SROB=summary risk of bias; RR=risk ratio (relative risk); MD=mean difference; OR=odds ratio; SD=standard deviation; 95% CI=95 percent confidence interval; NR=not reported; SS=statistically significant; NS=statistically not significant; mo=month(s); yr(s)=year(s)

Only those outcomes listed for which any evidence was reported; studies not reporting a given outcome are not listed

* Favours THR (or RS), no difference, or inconclusive

** Decision was consensus-based

[£] Calculated

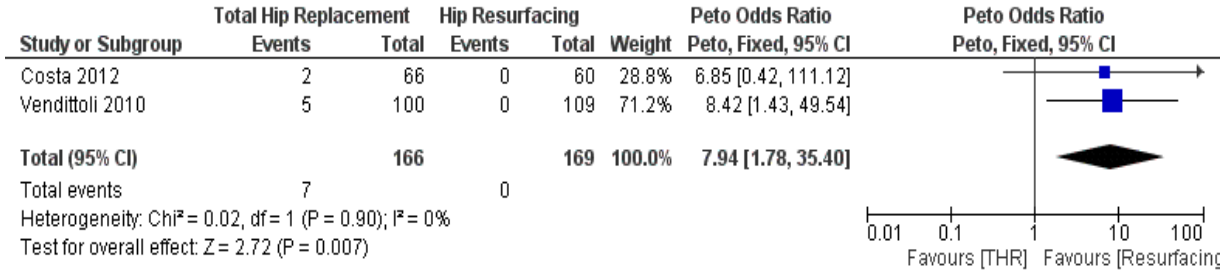


Figure 15. Risk of infection

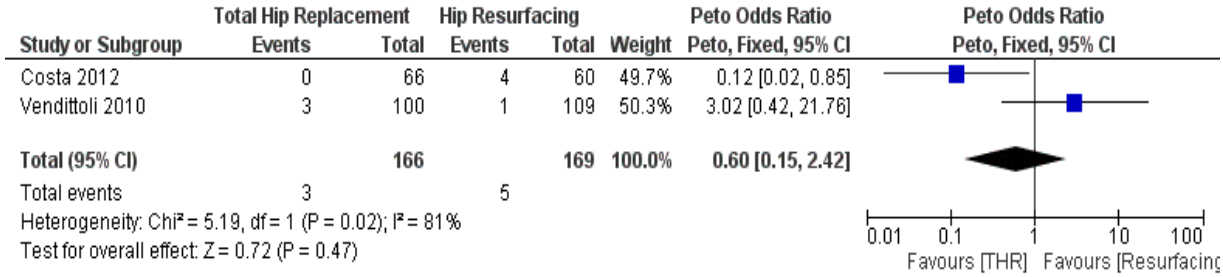


Figure 16. Risk of deep vein thrombosis

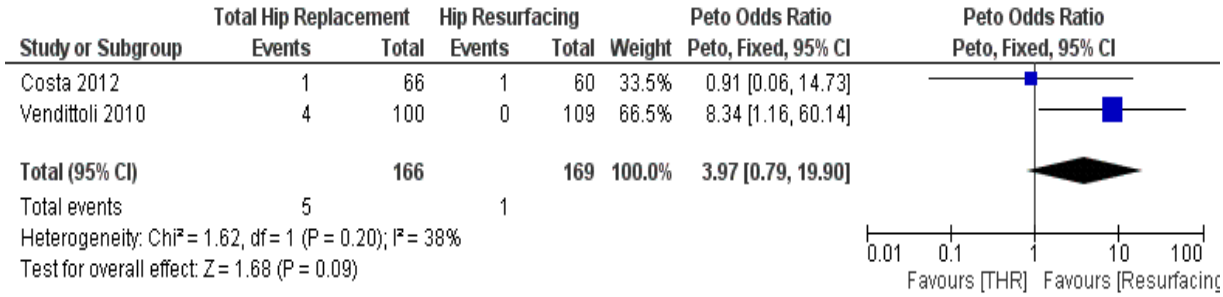


Figure 17. Risk of implant dislocation

Evidence from systematic reviews

Functional/clinical measures

Two of the three included systematic reviews comparing THR to RS, reported evidence on HHS (see Table 42),[139,140] WOMAC,[139,140] Merle d'Aubigne and Postel,[139] and UCLA scores.[139] The evidence was inconclusive due to the lack of pooled mean difference estimates for all three scores as well as the inconsistent results for the mean Harris Hip and WOMAC scores.

Health related quality of life

No evidence was identified.

Revision

Both systematic reviews found a higher risk of revision in patients receiving RS as compared to THR.[139,140] One review data meta-analysed from four RCTs which compared risk of revision in RS- vs. THR-recipients and reported a pooled RR estimate of 2.60 (95% CI: 1.31, 5.15) (see Table 42).[139]

Mortality

Overall, evidence on mortality reported by the two systematic reviews was inconclusive due to great uncertainty in the effect estimates and the variability around them.[139,140] For example, the pooled RR for mortality compared between RS and THR and reported in one review was 1.10 (95% CI: 0.10, 17.8) (see Table 42).[140]

Failure rate

One systematic review reported an indirect naïve 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).[141] The analysis suggested a reduced risk of failure in the RS vs. THR recipients (3.70% vs. 11.60%). Given the well-recognized problems with validity of such methodology, the evidence was judged to be inconclusive by us.

Complications

Evidence on complications was reported for both systematic reviews (i.e., implant dislocation, infection, and component loosening) (see Table 42).[139,140] The evidence consistently showed an increased risk for component loosening,[139,140] but reduced risk for implant dislocation[139] amongst RS recipients compared to THR recipients. One review,[139] 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, 8.31).

Table 42. Summary results for resurfacing arthroplasty vs. total hip replacement – Systematic reviews

Follow-up	Pooled effect estimate (95% CI) [RS vs. THR]	# of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect Conclusion*
Harris Hip score (range: 0-100)				
1-2 yrs	NR[139] No significant difference	3[139]	Medium quality[139]	Inconclusive
2 yrs	MD=2.51, 95% CI: 1.24, 3.77 (SS)[140] Better in RS vs. THR	NR[140]	High quality[140]	
Western Ontario and McMaster University Osteoarthritis Index score (range: 0-100)				
1-2 yrs	NR[139] No significant difference	3[139]	Medium quality[139]	Inconclusive
2 yrs	MD=-2.41, 95% CI: -3.88, -0.94 (SS) [140] Better in HRA vs. THR	NR[140]	High quality[140]	
Merle d'Aubigne and Postel score (range: 0-18)				
1-2 yrs	NR[139] No significant difference	3[139]	Medium quality[139]	Inconclusive
University of California, Los Angeles activity score (range: 1-10)				
1-2 yrs	NR[139] The mean UCLA activity scores significantly higher in RS vs. THR	2[139]	Medium quality[139]	Inconclusive
Revision rate (n/N)				
1-10 yrs	RR=2.60, 95% CI: 1.31, 5.15 (SS)[139]	4[139]	Medium quality[139]	In favour of THR
NR	RR=1.72, 95% CI: 1.20, 2.45 (SS)[140] Higher in RS vs. THR (19 pooled RCTs and non RCTs)	NR[140]	High quality[140]	
Mortality rate (n/N)				
3 yrs	NR[139] One study showed no significant difference between RS vs. THR RR=1.05, 95% CI: 0.24, 4.66	1[139]	Medium quality[139]	Inconclusive
NR	RR=1.10, 95% CI: 0.10, 17.8 (NS)[140]	NR[140]	High quality[140]	
Failure rate (n/N)				
NR	3.70% (95% CI: 2.0, 6.5) vs. 11.60% (95% CI: 7.50,17.40)[141] Indirect naïve comparison of 15 studies of RS and 19 studies of THR	NA[141]	Low quality[141]	Inconclusive
Dislocation rate (n/N)				
1-2 yrs	RR=0.25, 95% CI: 0.05, 1.21(NS)[139]	3[139]	Medium quality[139]	In favour of RS
NR	RR=0.20, 95% CI: 0.10, 0.50 (SS)[140] Lower in RS vs. THR (#pooled studies NR)	NR[140]	High quality[140]	
Component loosening (n/N)				
1-10 yrs	RR=4.96, 95% CI: 1.82, 13.50(SS)[139] Higher in RS vs. THR	4[139]	Medium quality[139]	In favour of THR
NR	RR=3.00, 95% CI: 1.11, 8.50 (SS)[140]	NR[140]	High quality[140]	

	Higher in RS vs. THR (10 pooled RCTs and non RCTs)			
Infection (n/N)				
1-3 yrs	RR=2.25, 95% CI: 0.61, 8.31(NS)[139]	3[139]	Medium quality[139]	Inconclusive
THR=total hip replacement; RS=resurfacing arthroplasty MD=mean difference; 95% CI=95 percent confidence interval; NR=not reported; yr(s)=year(s); PE=polyethylene; SS=statistically significant; NS=statistically not significant; MA=meta-analysis				

Only those reviews listed for which any evidence for the given outcome was reported
Favours THR (or RS), no difference, or inconclusive

6.2.6 Grading overall quality of evidence

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

Table 43. GRADE evidence profile for gradable outcomes reported in RCTs of THR vs. RS (adapted from Guyatt 2011[99])

Outcome [follow-up timing]	N 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)*
THR vs. RS – 3 RCTs[127-129]								
Harris Hip score [12 mo]	1 (126)[127]	None **Inconclusive**	Low	NA	Direct	Imprecise	Likely	Very low
WOMAC score [3-24 mo]	2 (313)[128,129]	None **No difference**	Unclear	Consistent	Direct	Precise	Likely	Low
Revision [3-56 mo]	1 (209)[129]	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 mo]	2 (335)[127,129]	OR=3.97 95% CI: 0.79, 19.90 **Inconclusive**	Low	Inconsistent	Direct	Imprecise	Likely	Very low

GRADE= Grading of Recommendations, Assessment, Development, and Evaluation; RCT=randomised controlled trial; CI=confidence interval; SROB=summary risk of bias; RCT=randomised controlled trial; NA=not applicable; yr(s)=year(s); mo(s)=month(s); RS=resurfacing; THR=total hip replacement

*GRADE categories: high, moderate, low, very low, NA (no evidence)

6.3 Summary conclusions comparing THR and RS

In three RCTs[127-129] and three systematic reviews[139-141] (Table 44 and Table 45) comparing THR and RS the majority of evidence was rated as inconclusive (RCTs - very low grade evidence). Nevertheless, the evidence from two RCTs and two systematic reviews indicated reduced risk of infection[127,129] and implant dislocation[139,140] amongst RS patients compared to THR patients. However, the evidence from the same reviews also indicated that recipients of RS were at higher risk of revision and component loosening compared to patients who received THR. In three RCTs,[127-129] mean post-operative OHS, WOMAC (low grade evidence), and Merle D'Aubigne Postel scores were not different between patients who received THR and RS.

There was inconclusive evidence on mortality (3 RCTs[127-129] and 2 systematic reviews[139,140]), HHS (1 RCT[127] and 2 systematic reviews[139,140]), UCLA score (2 RCTs[128,129] and 1 systematic review[139]), and selected complications (i.e., infection, wound complication, deep vein thrombosis) in two RCTs[127,129] and one systematic review.[139]

Results from individual RCTs were considered inconclusive due to the partial reporting (missing data for effect estimates, confidence intervals, standard errors, standard deviations, p-values) and great uncertainty in the estimates (wide CIs). The findings from systematic reviews were inconclusive due either to great uncertainty in the pooled estimates (wide CIs), unreported pooled results across RCTs (i.e., reported only narrative synthesis), or reporting inconsistent summary findings.

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

Conclusive evidence suggesting difference	Conclusive evidence suggesting no difference	Inconclusive evidence
RCTs (THR vs. RS)[127-129]		
Infection[127,129] **In favour of RS**	Oxford Hip score[127] WOMAC score [low grade evidence][128,129] Merle D'Aubigne and Postel score[129]	Harris Hip score [very low grade evidence][127] UCLA score[128,129] SF-36[128] Euro-Qol EQ-5D[127] Revision [very low grade evidence][129] Mortality [no evidence-not graded] Deep vein thrombosis[127,129] Implant dislocation [very low grade evidence][127,129] Superficial wound complication[127] Aseptic loosening[129]
RCT=randomised controlled trial; total THR=total hip replacement; RS=resurfacing; SF-36=Short Form Health Survey; UCLA= University of California, Los Angeles activity scale; WOMAC=Western Ontario and McMaster University Osteoarthritis Index		

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

Conclusive evidence suggesting difference	Conclusive evidence suggesting no difference	Inconclusive evidence
Systematic reviews (THR vs. RS)[139-141]		
Revision[139,140] **In favour of THR** Implant dislocation[139,140] **In favour of RS** Component loosening[139,140] **In favour of THR**	None	Harris Hip score[139,140] WOMAC score[139,140] Merle D'Aubigne and Postel score[139] UCLA score[139] Mortality[139,140] Failure[141] Infection[139]
RCT=randomised controlled trial; total THR=total hip replacement; RS=resurfacing; UCLA= University of California, Los Angeles activity scale; WOMAC=Western Ontario and McMaster University Osteoarthritis Index		

6.3.1.1 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 due to insufficient numbers of data points in the forest/funnel plots.

Heterogeneity, subgroup effects, and sensitivity analysis

The reviewed data from RCTs was too sparse (only 3 RCTs) to allow the exploration 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 characteristics (i.e., age, sex, or functional status). None of the included RCTs reported within-study subgroup effects of the THR relative to RS (or vice versa).

6.4 Overall summary of clinical effectiveness findings

A large proportion of evidence appraised and summarized in this review has been inconclusive (very low to low grade) due to poor reporting, missing data, inconsistent results, and/or great uncertainty in the treatment effect estimates. Notwithstanding, results from most studies suggested significantly improved post-surgery scores for functional/clinical measures (HHS, OHS, WOMAC, MACTAR, Merle D'Aubigne Postel, and SF-12) in participants, regardless of the type of THR or RS they received. Some moderate or lower grade evidence indicated the absence of difference for these measures between types of THR (or between THR vs. RS) at different follow-up times. There was a reduced risk of implant dislocation for participants receiving THR with larger femoral head size (vs. smaller head size) or with cemented cup (vs. cementless; high grade evidence). Moreover, the evidence suggested reduced femoral head penetration rate (moderate grade) and risk of implant revision (very low grade) for participants who received cross-linked polyethylene vs. conventional polyethylene cup liner bearings. Participants with ceramic-on-ceramic articulations (vs. metal-on-

polyethylene) experienced reduced risk of osteolysis. Recipients of RS, compared to THR, had a lower risk of infection. The evidence on mortality and other complications (e.g., loosening, femoral fracture, and deep vein thrombosis) was inconclusive (very low grade).

Limitations of reviewed evidence and pitfalls in interpretation

The review findings warrant cautious interpretation given the limitations of the reviewed evidence. Specifically, great uncertainty in the treatment effect estimates (i.e., wide 95% confidence intervals) due to limited sample size and/or small numbers of events (especially for deaths, revisions, and complications), as well as incomplete or poor reporting (e.g., missing effect measures, standard deviations/errors, 95% confidence intervals, p-values) rendered some of the reviewed evidence inconclusive. Moreover, reported evidence on complications was scarce. It is unclear whether this is due to the absence or rarity of these events or it is simply due to underreporting. In light of poor reporting, it was not possible to explore contextual factors which 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 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 whether there was any variation in the treatment effect across the pre-defined subgroups of patients or methodological features of studies; likewise, the extent of publication bias could not be examined via funnel plots due to the small number 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 vs. THR and THR vs. RS comparisons trials were often based on selective populations or interventions. Studies and provided data on revision rates which were inconclusive with often wide confidence intervals.

Comparison of results from RCTs and SRs

The findings of RCTs and SRs could be compared only with regards 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 SRs were statistically non-significant and had wide uninformative confidence intervals around them. Therefore, the evidence from both RCTs and SRs was rendered as inconclusive due to wide variability around the estimates and/or missing data. The reviewed evidence from RCTs suggested that there was no difference in post-operative Harris Hip scores between cemented and cementless THR groups. The evidence for Harris Hip score reported in the included SRs was ruled as inconclusive.

Our update search identified four new relevant SRs.[239-242] Of these four SRs, three compared the effectiveness of THRs using different articulations (metal-on-metal vs. metal-on-polyethylene),[239] implant fixation methods (cemented vs. cementless),[242] or femoral stem coating materials (hydroxyapatite-coated vs. non-hydroxyapatite-coated)[241] for risk of revision,[242] Harris Hips score,[239,241,242] mortality,[242] and complications.[239,242] One remaining SR compared THR to RS for risk of revision.[240]

Briefly, the review by Voleti et al.[239] presented a meta-analysis based on three RCTs and found no significant difference for Harris Hip score between the two articulations (metal-on-metal vs. metal-on-polyethylene) at 6 years of post-surgery follow-up (pooled MD= -1.05, p=0.37). However, the risk of complication (dislocation, aseptic loosening, trochanteric/iliopsoas bursitis, femoral fracture, and wound dehiscence) was greater in the metal-on-metal compared to the metal-on-polyethylene articulation group (OR=3.37, 95% CI: 1.57, 7.26).[239] Similarly, another review[242] presented a meta-analysis of seven RCTs showing statistically non-significant difference in the mean post-operative Harris Hip score between the cemented and cementless THR groups (pooled MD= 1.12, 95% CI: -1.17, 3.41). In the same review, the meta-analytic estimates for risk of revision (6 RCTs; pooled RR=1.44, 95% CI: 0.88, 2.36), mortality (5 RCTs; pooled RR=1.06, 95% CI: 0.73, 1.52), and complications (4 RCTs; pooled RR=1.54, 95% CI: 0.21, 11.03) between cemented vs. cementless groups of THR were also statistically non-significant. In the review by Li et al.,[241] the post-operative pooled mean Harris Hip score was not statistically significantly different between the hydroxyapatite-coated vs. non-hydroxyapatite-coated THR groups (4 RCTs; pooled MD= 3.04, 95% CI: -4.47, 10.54). The review by Pailhe et al., 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.[240]

In summary, the findings from the newly identified SR[242] are in agreement with those of this review in showing no difference in post-operative Harris Hip scores between cemented vs. cementless THR groups. Also in agreement with our findings, the pooled estimates for revision, mortality, and complications were statistically non-significant with sufficiently wide 95% CIs (due to low event

counts and small sample size of trials) that were compatible with moderate-to-large effect size in either direction, rendering these findings as inconclusive.[242] Future well-designed randomized trials need to corroborate or refute the finding of one SR[239] suggesting increased risk of complications in the metal-on-metal vs. metal-on-polyethylene articulation group.

Strengths and limitations of the review

One of the strengths of this review is based on 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 limitations of the evidence itself, the scope of this review was limited to a pre-defined set of outcomes ascertained from recently published evidence (2008 or later); evidence from studies with sample size less than 100 participants and non-English publications was not included. Given the wide scope and large amount of identified evidence, we limited our inclusion to studies with a sample size of 100 that were published since 2008. The rationale for such limitations was based on the fact that smaller studies tend to be underpowered to detect meaningful differences in outcomes.[243,244]

The results of such studies are usually rendered as inconclusive due to statistically non-significant estimates with wide confidence intervals that include large treatment effect size values compatible with both a better and worse outcome for any given treatment compared to the control treatment. Therefore, in order to minimize this problem, we calculated the minimum sample size for a study which would have 90% power, at two-tailed test significance level of 0.05, to detect the mean difference of 10 on HHS (we selected a standard deviation of 15 based on external sources).[104,245] This calculation yielded a total sample size of 100 participants.

Future research

Since the evidence for any given comparison of two THRs was sparse (maximum of two trials), observed findings need to be replicated in new larger long-term pragmatic trials comparing the same THRs to each other (or RS) before more definitive conclusions or recommendation are made. The conduct of large multi-centre long-term pragmatic trials would help to evaluate relative treatment effects and their variation(s) across patient- as well as manufacturer-based subgroups more reliably and maximize generalizability of the findings to larger populations in clinical practice settings. For a more complete picture to aid health care policy decisions, new trials are also needed to consider the measurement of cost-effectiveness of alternative hip replacement (or RS) techniques. Study authors are encouraged to specify minimal clinically important differences and power calculations for their primary outcome(s). This information would help to interpret the study findings both in terms of clinical and statistical terms.

Better reporting of future trial results is warranted.

6.5 Cost-effectiveness methods

6.5.1 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 Appendices 1 & 2). An iterative procedure was used to develop these strategies with input from clinical advisors and previous HTA reports (e.g. Vale et al., 2002;[19] deVerteuil et al., 2008[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.[246,247]

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 2).

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:

- Searching of electronic bibliographic databases, including research in progress
- Scrutiny of 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; The Cochrane Library (specifically Cochrane Database of Systematic Reviews, CENTRAL, DARE, NHS EED, HTA database); and CEA Registry (Articles).

The following databases of research in progress were searched: Current Controlled Trials; ClinicalTrials.gov; UKCRN Portfolio Database; and NLM Gateway (Health Services Research Projects in Progress (HSRProj)).

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

6.5.2 Inclusion criteria

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

Study design

- Randomised controlled trials
- 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 arthroplasty

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 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 or QALY); ICER, uncertainty measures, the ceiling willingness-to-pay ratios, and probabilities from cost-effectiveness acceptability curves.

Exclusion criteria

- Non-English language publications

- Abstract/conference proceedings, letters, and commentaries
- Quality of life reported without utility or QALY
- Hip/knee data not reported separately
- Studies including only patients <35 years

6.5.3 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
- Animals
- Children
- Surgery due to hip fracture only
- Non English full-text

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.[96]

6.5.4 Data extraction

Data extraction was carried out in two stages by one reviewer using the data extraction sheets (see Appendices 11-13) and checked by a second reviewer. Stage one considered all eligible studies and stage two considered studies assessed for usefulness to populate the economic model. Stage one data extraction included the following:

- Study characteristics (i.e., author's name, country, design, study aim, type of economic evaluation (i.e., cost-effectiveness, cost-utility analysis), perspective (e.g., societal, health care payer, patient) and study currency)
- Patient characteristics (i.e. number of participants, age, gender, osteoarthritis)
- Outcomes (i.e. utilities, resources use and cost (both direct and indirect), incremental cost-effectiveness ratios)

Data extraction also included the overall study conclusion and a comment on the type of data included in the studies 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:

Cost:

1. UK study ≥2008
2. UK study < 2008
3. Non-UK study ≥2008
4. Non-UK study < 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:

Utilities:

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

Second stage data extraction considered cost of THR (cost of device, cost of surgical time/ cost of hospital stay), cost of follow up for successful THR, revision THR, follow-up for successful revision THR, costs of RS (cost of device, cost of surgical time/ cost of hospital stay), costs of 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. Information on definition of costs, source of costs, cost year and currency was also extracted.

6.5.5 Quality assessment

The key cost-effectiveness papers which 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) [248] list, while cost effectiveness studies with economic models were also assessed using the Philips criteria.[249]

6.6 Cost-effectiveness results

6.6.1 Identification of studies

The flow chart outlining the process of identifying relevant literature can be found in Figure 18. The database search on cost-effectiveness identified 1650 records with an additional 14 records identified through screening of reference lists of included studies. Duplicate removal left 913 studies to be screened for inclusion. The initial sift excluded 283 studies that were clearly not relevant and a

further 525 records were excluded on title and abstract (kappa score: 0.89). 105 full texts were assessed for eligibility of which 35 were excluded with reason (Appendix 14). This resulted in a total of 70 eligible records including 66 studies that were subsequently included in the review. Of these 35 were observational studies with or without economic analysis, 22 were economic analyses including three HTAs, four reviews (3 non-systematic and one systematic), four RCTs and one was a before and after trial. Study location covered the UK (n=13), other European countries (n=22), North America (n=21), Australia and New Zealand (n=6) and Asia (n=4). Cost/resources were reported by 30 studies, utilities/QALYs by 15 studies and 21 studies reported both, cost/resource use and utilities/QALYs. Seven of the 14 economic models reported transition probabilities.

A separate search (December 2012) of the Clinical Trials.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 to be potentially relevant from the cost-effectiveness point of view (Appendix 8). All were either on-going or completed since 2009.

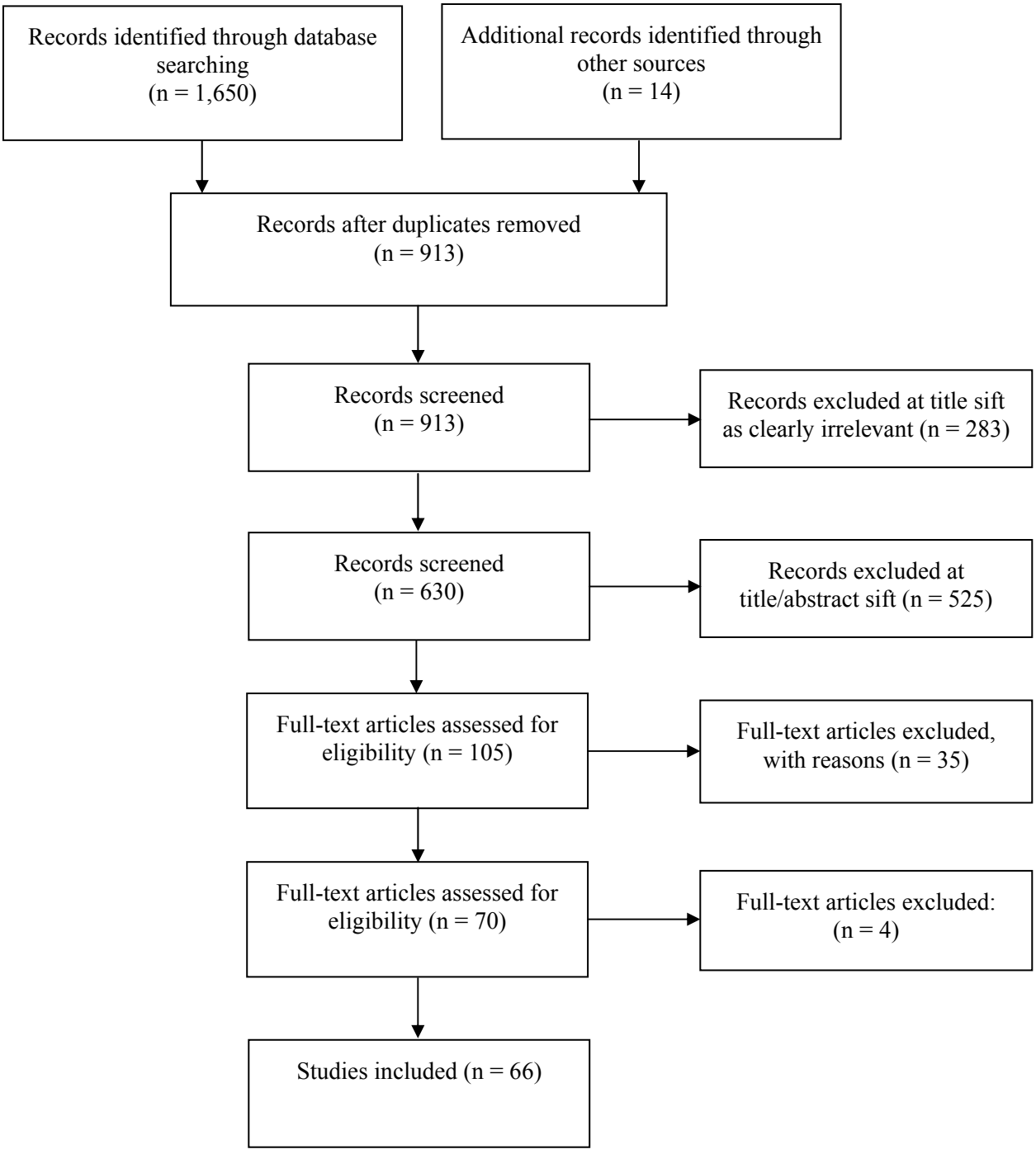


Figure 18. PRISMA flow diagram cost-effectiveness studies

6.6.2 Description of included studies

6.6.2.1 RS arthroplasty

Evidence on RS was scarce with only five of the 66 included studies investigating hip RS arthroplasty (Appendix 12). A 2012 UK RCT investigated the cost-effectiveness of RS compared with THR including 126 OA patients suitable for RS.[40,127] At the end of this 12-month trial small benefits in terms of QALYs for RS could be shown for a selected patient group resulting in an ICER of £17,451 per QALY. This evidence was stronger for male than female patients. In a comparison of ceramic-on-ceramic THR with RS three-months post-surgery evidence was not as strong favouring THR over RS. [204] However, longer-term follow up in a study comparing hybrid THR with RS confirmed that after five and nine 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.[250,251]

A retrospective economic analysis of published data over a 30-year time horizon showed cost-effectiveness of RS compared to THR for women <55 and men <65 years of age.[252] The main drivers of cost-effectiveness were cost of implant and length of hospital stay.[40,204] However, Vale et al. (2003) reported in their HTA that RS compared to THR would only be cost-effective if revision rates could be shown to be 80-88% lower than revision rates for THR.[19] They further concluded that RS could be cost-effective when compared with 'watchful waiting' followed by THR or when compared with an extended period of 'watchful waiting' over 20 years.

6.6.2.2 THR

The majority of studies investigated THR (n=61) (Appendix 11). 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.[253 11,254,255]. However, Coyle et al. (2008) concluded that there is little evidence of a difference between the two surgical techniques in the long-term mainly due to lack of data.[144]

Ten of the THR studies focused on the comparison of different types of THR or specific components/brands of THR. Briggs et al. (2004), Davies et al. (2010), Fordham et al. (2012) and Hulleberg et al. (2008) assessed different brands of THR,[43,256,257] Bozic et al. (2003) investigated alternative bearings including metal-on metal, ceramic-on-ceramic and polyethylene [258] and Laupacis et al. (2002), Marinelli et al. (2008), Pennington et al. (2013) and di Tanna et al. (2011) compared cemented, cementless and hybrid THR more generally and reported inconsistent findings.[44,117,259,260] The most recent economic model by Pennington et al. (2013)[44] used patient reported outcome measures 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. (2010)

identified cemented prostheses as the least costly 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.[43] Cummins et al. (2009) reported that use of antibiotic impregnated bone cement can result in an overall cost decrease.[261] For more detail on the studies investigating the different types of THR see Appendix 13.

Patient management and rehabilitation was the focus of four studies which reported that perioperative management and rehabilitation programmes could improve patient outcomes and reduce costs.[262-265]

The majority of the THR studies (34/61) assessed the cost and/or effectiveness of THR without specific focus on a rehabilitation programme, surgical intervention, implant brand or prosthesis type. Of these, two US studies 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.[266,267] Eleven studies evaluated the cost-effectiveness of THR for a specific country,[268-278] while two multicentre studies aimed to assess the costs and outcomes of THR comparatively across a number of European member states.[279,280] These two studies concluded that improvement after surgery is associated with high pre-operative expectations. Stargardt et al. (2008) reported further that total cost of treatment ranged from €1,290 (Hungary) to €8,739 (The Netherlands) and that the two main cost drivers were the cost of the implants and ward costs.[279]

Overall findings of the cost-effectiveness studies were that 1) THR resulted in greater benefits when compared to conservative treatment and 2) waiting times incurred greater costs and resulted in physical deterioration.[270,281,282] Further agreement was reached on the long-term cost-effectiveness and sustained benefits for THR. [37,117,256,272,274] However, Bozic et al. (2011) stated that while THR improved quality of life, failed THR could lead to health states worse than chronic OA.[283] Resource use might be increased as patients with THR were shown to have a 10% increase in hospital stay compared to pre-surgery.[284]

In contrast, two studies that took a patient perspective rather than a health care perspective concluded that out-of-pocket costs (including hospital costs, medications, rehabilitation costs, health professional visits, tests, special equipment, household alterations, use of private and community services, transportation costs that are not paid for by 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.[285,286]

Studies that focused on revision THR concluded that revision THR seemed cost-effective but that they were resource intensive and have important implications for the allocation of healthcare funding

since the number of revisions is expected to increase with increasing demand for THR.[287-289 290] Vanhegan et al. (2012) 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 National Health Service tariffs.[291] Durable implants and reduction in complications such as early dislocations have been suggested to be the solutions to reduce revision rates.[288] However, highest revision costs were reported for revision due to infection[291] and that infections due to methicillin-resistant strains of bacteria (41% of peri-prosthetic joint infections) incurred significantly higher costs than revision of infections of sensitive strains.[292]

Four studies evaluated the usefulness of different outcome measures for quality of life after THR or revision THR which showed that there was no consistency in tools used to assess quality of life. Feeny et al. (2004) reported that there is low agreement between certain outcome measures (SF-36, standard gamble, HUI-2 and HUI-3).[293] Dawson et al. (2001) and Jones et al. (2012) found that disease-specific measures reported larger changes than generic and utility measures.[294,295] Ostendorf et al. (2004) 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.[296]

Overall, studies confirmed long-standing claims that THR and RS are cost-effective interventions for patients with osteoarthritis of the hip. However, there is little evidence from long-term trials on the comparison of different implant brands and types of prostheses allowing no conclusions to be drawn on the most cost-effective prostheses type let alone specific brands within the types or specific patient groups which might benefit most from surgery. Studies used different methodologies to estimate costs (reference costs vs. prices actually paid by health care centres) and definitions of costs included varied extensively while many studies did not clearly report how costs involved were broken down. While 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 focus questions on those needed to support cost-effectiveness studies to inform future resource allocation.

6.6.2.3 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 vs. THR,[252] the

cost-effectiveness of cemented, cementless and hybrid prosthesis[44] and 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. (2012) reported costs for revision THR.[291] Baker et al. (2011)[251] and Hulleberg et al. (2008)[257] reported medium to long term utilities in small populations, Dawson et al. (2001)[294] investigated quality of life post revision THR and Bozic et al. (2011)[283] 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. (2011) evaluated the Swedish PROMs data reporting utilities for close to 35,000 THR patients.[297]

Of these 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.[19,44,291] (see Appendix 13) Costs for RS was reported in three studies.[19,40,252] Of these Edlin et al. (2012)[40] and Vale et al. (2002)[19] also reported follow-up costs after successful RS and Bozic et al. (2010) reported costs for revision RS[252] (Appendix 12).

The studies identified to report the most useful data on utilities following THR were Pennington et al. (2013), Rolfson et al. (2011), Hulleberg et al. (2008), Dawson et al. (2001) and Bozic et al. (2011)[44,257,283,294,297] (see Appendix 15). Utilities for RS were only reported in three studies[40,251,252] (see Appendix 16). No data were identified on quality of life >12 months post-RS or for post-revision RS. Follow-up costs reported by Vale et al. (2002) were the same for THR, RS and revision THR.[19] Similarly, Bozic et al. (2010) made no distinction between revision following THR and RS in terms of costs.[252]

6.6.2.4 Quality assessment of core studies

Of the 11 core studies, five studies (Baker et al., 2011; Hulleberg et al., 2008; Rolfson et al., 2011; Dawson et al., 2001; Bozic et al., 2010)[251,252,257,294,297] had useful information only on EQ-5D utility scores and one study (Vanhegan et al., 2012)[291] provided useful data on costs only. These partial economic evaluations were not included in the critical appraisal.[298]

Five studies were full economic evaluations and have been critically appraised using the CHEC-list.[248] Of these five studies, four included models. These studies have also been critically appraised using an adapted checklist for models developed by Philips et al (2006). [249]

Table 46 shows that all studies met 16 or more of the 19 criteria for economic analyses given by Evers et al. (2005).[248]

Table **47** shows that all studies met 20 or more of the 32 criteria for economic models given by Philips et al. (2006).[249] All studies had correctly reported the time horizon and the perspective of the model, and the inputs used within the model were consistent with the perspective which was chosen. In terms of costs and outcomes used in the model these were appropriate to the specific study dataset which 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. (2013)[44] did not provide a clear definition of all the options under evaluation and Briggs et al. (2004)[38] did not specify the cycle length of the model.

Table 46 Critical appraisal of the economic evaluation studies using the CHEC-list[248]

CHEC-List	Bozic et al. (2010) [252]	Briggs et al. (2004) [38]	Edlin et al. (2012) [40]	Pennington et al. (2013) [44]	Vale et al. (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	N/A	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 generalizability 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

Key: Y = yes, No = no, UN = unclear, N/A = not applicable

Table 47 Critical appraisal of the economic models using an adapted Philips checklist [249]

Philips criteria		Bozic et al. (2010) [252]	Briggs et al. (2004) [38]	Pennington et al. (2013) [44]	Vale et al. (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?	N/A	N/A	N/A	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	Y	Y	Y	Y

	justified?				
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
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	N/A
30	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	N/A	N/A	N/A	N
31	Has heterogeneity been dealt with by running the model separately for different sub-groups?	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

Key: Y = yes, No = no, UN = unclear, N/A = not applicable

6.6.2.5 Core studies for the economic model

Out of the 11 core studies, Edlin et al. (2012), Pennington et al. (2013), Vale et al. (2002) and Vanhegan et al. (2012) provided data for the model in Chapter 10.[19,40,44,291] Please see Chapter 10 for the rationale of the selection procedure. This section will provide a brief description of the four core studies in text and in Table 48.

Edlin et al. (2012) reported a cost-utility analysis of RS versus THR of a randomised controlled trial of 126 adult patients with severe arthritis of the hip.[40] Patients were randomised on a 1:1 basis between THR and RS. All RS patients received a Cormet MoM 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 months trial period. The costs were reported as 2009/2010 figures and 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 individual patient data 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 if the cheapest THR type (MoM) was used 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 initial operation/care, subsequent inpatient, outpatient, primary and community care, aids and medications (THR: £7,217 (SD: 1,320) and RS: £6,653 (SD: 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 a reported 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 variations of costs within each type of prostheses. Variation in prostheses costs by hospital, a change in current practice regarding the choice of THR implant, longer follow up (including higher revision rates for RS compared to THR) and use of different RS implants may affect the reported cost-effectiveness in this study.

Pennington et al. (2013) used individual patient data from three data sources (the national patient reported outcome measures programme, the NJR for England and Wales and hospital episode statistics) to compare the cost-effectiveness of cemented, cementless and hybrid THR in adult patients with hip osteoarthritis.[44] 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 adjusted for LOS by prosthesis type and costs of revisions varied by reason for revision. Costs were reported as 2010-11 prices. The national data sources provided data for 30,203 patients on QOL, LOS, rates of revision and re-revision and mortality.

Patients receiving different prosthesis types were matched by age, gender, number of comorbidities and ASA, BMI and deprivation, pre-operative QOL, surgeon experience and hospital type. The study reported data on combined costs of prostheses, operating theatre and hospital, QOL at six months post-surgery as well as five- and 10-year revision rates by prosthesis type, age group and gender. Overall the study concluded that in patients aged 70 years the ICER for hybrid prostheses compared to cemented prosthesis was £2,100 for men and £2,500 for women with hybrid prostheses resulting in higher QOL in all subgroups except women aged 80 and cemented prostheses being the least costly option. Initial costs of cementless prostheses were highest in all subgroups. One of the limitations of the study was that it assumed that the observed QOL at six months post-surgery would remain unchanged for the patients' lifetime. Furthermore, the study did not consider different revision rates by brand within the three different THR types.

Vale et al. (2002) undertook a HTA of the effectiveness and cost-effectiveness of RS compared to watchful waiting (i.e. patient monitoring, drug-based treatment and supportive activities including physiotherapy), THR and other bone conserving treatments.[19] The HTA comprised a systematic review of the clinical and cost-effectiveness of RS compared to 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 2000/01 £) for THR and revision THR were taken from the literature (£4,195 and £6,027 respectively) and prostheses costs for RS were obtained from manufacturers. The model considered the lower of the two obtained RS implant costs (£1,730 versus £1,890) resulting in

an overall cost of £5,515 for RS. LOS was estimated to be 10 or 12 days for THR and eight or 10 days for RS. All other costs including use of operating theatre, staff, x-rays, 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 X-ray totalling to £118.74. QOL estimates considered pain levels and QOL scores for mild, moderate and severe OA and were combined with revision and mortality rates to generate QALYs.

The main conclusion from the systematic reviews was that evidence from the literature on the effectiveness of RS was limited. Revision rates were reported to range between 0-14% over a three-year follow-up period for RS compared to revision rates of 10% or less 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 less than 80-88% of 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 due to 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. (2012) investigated the costs of 305 consecutive revision THR by reason for revision in 286 patients with a diagnosis of hip OA in 64% of revisions (n=195).[291] 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 2007/08 £) and included costs of implant, materials and augmentation, use of the operating theatre and recovery room, the inpatient stay and costs of 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, peri-prosthetic fracture and dislocation).

The mean cost of aseptic revision was reported to be £11,897 (SD: 4,629), £21,937 (SD: 10,965) for revisions for deep infection, £18,185 (SD; 9,124) for revisions of peri-prosthetic factures and 10,893 (SD: 5,476) for dislocations. Higher complication rates as well as re-operation rates were associated with revisions for deep infection, peri-prosthetic fracture and dislocation. However, numbers of revision for these three indications were relatively small (n=76, n=24 and n=11, respectively). While the cost estimates can be assumed to be very accurate their limitation is lack of generalisability since they were based on one single tertiary centre. Furthermore, the study did not consider cost of readmission for complications and other direct and indirect medical and social costs.

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

Author Year Country Study ID	Study design	Methods	Results	Main Conclusion	Information for model used
Edlin et al. (2012)[40] UK Costa et al. (2012)[127] UK	TYPE: RCT and cost-utility analysis AIM: To report on the relative cost-effectiveness of THR and RS arthroplasty in patients with severe arthritis suitable for hip joint RS arthroplasty	POPULATION: Patients >18 years with severe arthritis of hip joint suitable for RS (N=126): THR (N=66) and RS (N=60) OUTCOMES: primary: hip function (12 months post-surgery Oxford hip score and Harris hop score) secondary: QOL (EQ-5D), disability rating, physical activity level, complications, cost-effectiveness Incremental costs and QALYs ICER ECONOMIC ANALYSIS: NHS perspective 12-month time horizon Cost year: 2009/2010 (£) Univariate sensitivity analyses	Hip function: Mean Oxford hip score effect size: 2.23 (95% CI -1.52 to 5.98, p=0.070) Mean Harris hip score 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 versus 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, one year after receiving a THR versus RS RS arthroplasty appears to offer very short-term efficiency benefits over total hip arthroplasty within a selected patient group	1) a) Resource use b) Costs 2) a) Utilities b) QALYs 3) Transition probabilities
Pennington et al. (2013)[44] UK	TYPE: Retrospective cost-utility and decision analysis	POPULATION: Patients undergoing primary THR for OA (N=30,203 for QOL analysis) Male:	Lifetime costs: lowest with cemented prostheses	Cemented prostheses were the least costly type for THR For most patient groups	1) a) Resource use b) Costs 2) a) Utilities b) QALYs

	AIM: To evaluate the relative cost-effectiveness of cemented, cementless, and hybrid prostheses for elective THR surgery	cemented: 35.1% (N=4195) cementless: 44.6% (N=6548) hybrid: 38.0% (N=1350) Age (mean, SD): cemented: 72.4 (6.7) years cementless: 67.8 (7.2) years hybrid: 70.4 (7.2) years OUTCOMES: QOL 6 months post-surgery (Oxford Hip Score, EQ-5D) Lifetime cost effectiveness Costs (£) ICERs ECONOMIC MODEL: Health service perspective cost year: 2010/11 (£) Sensitivity analysis of QALY post 2 years, revision rates using different hazard function, failed hip category without revision, excluding metal on metal prostheses	postoperative QOL and lifetime QALYs: highest with hybrid prostheses Women aged 70: mean costs for cemented: £6,900 mean costs for cementless: £7,800 mean costs for hybrid: £7,500 mean postoperative EQ-5D scores: cemented:0.78 cementless: 0.80 hybrid: 0.81 lifetime QALYs: cemented: 9.0 years cementless: 9.2 years hybrid: 9.3 years ICER: hybrid vs cemented: £2,500/QALY	hybrid prostheses were the most cost effective Cementless prostheses did not provide sufficient improvement in health outcomes to justify their additional costs	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	
Vale et al. (2002)[19] UK McKenzie et al. (2003)[299] UK	TYPE: HTA systematic review and retrospective cost-utility analysis AIM: To assess the	POPULATION: Patients with hip disease Age: 45-50 and 65-70 years OUTCOMES:	Revisions: MoM over 3-year follow-up: 0–14% THR over 10-year follow-up: 10% or less Osteotomy over 10–17	MOM had lower revision rates than THR over an extended time period and resulted in better outcomes overall for persons who are likely	1) a) Resource use b) Costs 2) a) Utilities b) QALYs 3) Transition probabilities	

	effectiveness and cost-effectiveness of metal-on-metal hip RS arthroplasty compared with watchful waiting, THR, osteotomy, arthrodesis and arthroscopy of the hip joint	Incremental costs (£) and QALYs ICERs ECONOMIC MODEL: Markov Model 20-year time horizon NHS perspective Cost year: 2000 (£) Subgroup analysis considering those that would not outlive a THR Sensitivity analyses for revision rates, operation times, watchful waiting costs, time horizon and QOL	year follow-up: between 2.9% and 29% Patients pain free: MoM: 91% at 4 years THR: 84% at 11 years Arthrodesis: 22% at 8 years Costs: MoM for a patient aged <65 years: £5,515 THR: £4,195 Revision: £6,027 Arthroscopy: £951 Osteotomy: £2,731 Watchful waiting: £642 annually Cost-effectiveness: for patients <65 years MoM dominated by THR MoM dominated watchful waiting within a 20-year follow Incremental cost per QALY: MoM versus osteotomy: £3,039 MoM versus arthroscopy: £366	to outlive a primary THR If MoM has lower revision rates than THR over an extended period and results in better outcomes from subsequent THR, then MoM could possibly be considered cost-effective or even dominant	Comment: revision rates for MoM and THR, costs including prosthesis costs, broken down costs for watchful waiting

			For patients aged > 65 years, THR dominated MoM		
Vanhegan et al. (2012) [291] UK	TYPE: Retrospective economic analysis AIM: To evaluate the costs associated with revision THR for different indications	POPULATION: Patients with revision THR (N=286 and N=305 procedures) Male: Aseptic loosening (N=194): 34% (N=65) Deep infection (N=76):42% (N=32) Peri-prosthetic fracture (N=24): 25% (N=6) Dislocation (N=11): 28% (N=3) Age (mean, range): Aseptic loosening: 67 (20-89) years Deep infection:62 (29-83) years Peri-prosthetic fracture: 76 (31-88) years Dislocation: 79 (54-90) years OA: Aseptic loosening: 69% Deep infection: 48% Peri-prosthetic fracture: 80% Dislocation: 54% OUTCOMES: LOS Costs (£)	Mean total costs for revision surgery: Aseptic cases: £11,897 (SD 4,629) Septic revision: £21,937 (SD 10,965) Peri-prosthetic fracture: £18,185 (SD 9124) Dislocation: £10,893 (SD 5,476) Surgery for infection and peri-prosthetic fracture: Longer operating times, increased blood loss, increase in complications, longer LOS	Financial costs vary significantly by indication Variation is not reflected in current National Health Service tariffs	1) a) Resource use b) Costs 2) a) Utilities b) QALYs 3) Transition probabilities

(RCT-Randomised Controlled Trial; THR-total hip replacement; RS-hip resurfacing arthroplasty; QOL-quality of life, EQ-5D-European quality of life-5 dimensions; NHS-national health service; ICER-incremental cost effectiveness ratio; QALY-quality adjusted life years; OA-osteoarthritis, SD-standard deviation; HTA-health technology assessment; MoM-metal on metal

6.7 Summary of overall cost-effectiveness evidence

We found four of 11 studies from which we can source utility and cost data these were Edlin et al. (2012),[40] Pennington et al. (2013),[44] Vale et al. (2002)[19] and Vanhegan et al. (2012)[291] for the model. We assessed these using the checklists developed by Evers et al. (2005)[248] and Philips et al. (2006)[249] and we found them to be of varying quality. All studies met 16 or more of the 19 criteria for economic analyses given by Evers et al. (2005).[248] and all studies met 20 or more of the 32 criteria for economic models given by Philips et al. (2006).[249]

6.8 Registries methods

6.8.1 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 Appendices 1 & 2). The registry search strategy was designed to capture the generic terms for arthritis, THR and RS in addition to the word registry. Searches were not date limited for the registry search and were undertaken in November 2012 (see Appendix 2). 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; The Cochrane Library (specifically Cochrane Database of Systematic Reviews, CENTRAL, DARE, NHS EED, HTA database); and CEA Registry (Articles).

6.8.2 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, utility and outcome 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 arthroplasty

Comparator:

- Different types of primary THR compared surface replacement 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
- Less than 1,000 patients included in the registry study at time of publication
- Hip/knee data not reported separately

6.8.3 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
- Animals
- Children
- Non registry study
- Surgery due to hip fracture only
- Non English full-text

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 reviewers to identify eligible studies, this was cross checked by a second reviewer. Disagreement was resolved by a third reviewer. Reasons for exclusion of full text papers were documented.

6.8.4 Data extraction

Data extraction was carried out on the final 49 papers by one reviewer in two stages. Stage one considered all eligible studies and stage two was to ascertain the studies that would provide useful input to the economic model and survival analysis. Stage one data extraction included the following:

- Author surname
- Publication year
- Country of registry
- Year that registry data was collected
- Type of registry data collected
- Size of the registry database
- Description of the patient population
- Result of key outcomes

Data extraction of the overall aim of the paper and conclusion of the study was 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 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

6.9 Results of registry review

6.9.1 Identification of studies

The PRISMA flow diagram outlining the identification of registry studies can be found in Figure 19.[96] The database search on registry studies identified 541 publications with an additional one record identified through other sources. A total of 329 papers remained once duplicates were removed and these were screened for relevance. This process excluded a further 230 papers resulting in a final 99 papers that were that were screened at title and abstract level. A further 50 studies were excluded with a reason provided (Appendix 17).

A final total of 49 papers were included in final data extraction after stage one from ten countries worldwide: Japan (n = 1),[300] Australia (n = 5),[301-304 305] UK (n = 8),[306,306 15,16,307-309] Italy (n = 2),[260,310] Finland (n = 10),[311-320] Norway (n = 6),[321-325] USA (n = 5),[49,326-328] Denmark (n = 4),[329-332] Sweden (n = 3),[297,333,334] and Slovakia (n = 1).[335] In addition there were seven papers reporting outcomes from multinational registries.[336-342]

Following stage two exclusion (not most recent paper publication in a series or not most recent annual report from a national joint registry) of the final 49 papers 19 papers were removed. A final total of 30 papers were included in the narrative review to reflect the most recent publication in a series from each particular registry for both THR and RS. Please see Table 49 and Table 50 for a descriptive overview of these studies.

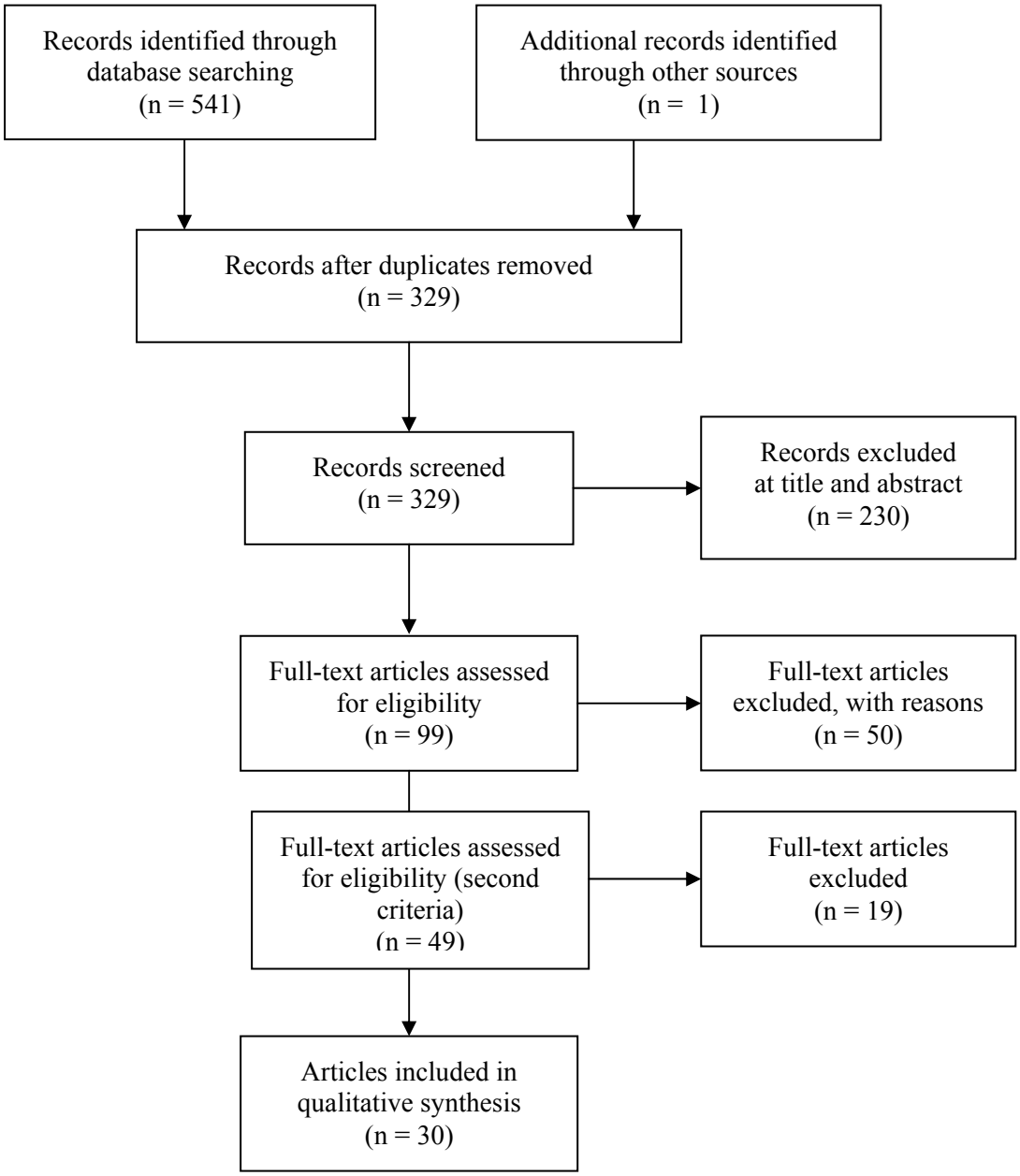


Figure 19 PRISMA flow diagram for registries studies

6.9.2 Review of included studies following stage two exclusion criteria

A narrative review of the included papers is given below by intervention (THR, RS) type and country. The 30 papers did not report similar patient populations, interventions, comparator groups or outcomes hence are reported separately. For the purposes of the economic model and survival analysis, revision rate or implant survival were the key outcomes to be extracted.

6.9.2.1 RS arthroplasty

Evidence on registry studies reporting RS arthroplasty was limited to eight studies, the majority of these studies investigated various comparison of THR surgery with RS. See Table 49 for a summary of the RS studies.

England and Wales

Jameson et al. (2012)[343] conducted a retrospective cohort study and reported survival time to revision for RS procedures from 2003-2013, it 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 (hazard ratio (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 (≤ 44 mm: HR = 2.14, 99% CI 1.53 to 3.00, p < 0.001, 45 to 47 mm: HR = 1.48, 99% CI 1.09 to 2.00, p = 0.001) than medium or large heads, as were operations performed by low volume surgeons (HR = 1.36, 99% CI 1.09 to 1.71, p < 0.001).

McMinn et al. (2012)[308] examined mortality and revision rates among patients with OA having THR, both cemented or uncemented procedures to compare against men undergoing RS. The authors used data from the NJR database for the analysis (154,996 receiving cemented THR, 120,017 receiving uncemented THR, and 8,352 receiving RS (in particular Birmingham hip resurfacing)). The baseline characteristics recorded include: age (cemented mean = 73.2, uncemented mean = 66.7), sex (cemented men = 53,409, women = 101,587 and uncemented men=50,529, women=69,488) and ASA grade. The analysis took into account the patient's age at primary surgery and their length of follow-up. Survival analysis was used to compare the cemented and uncemented with adjustment for sex, age at primary surgery, ASA grade before the operation, complexity of the procedure and 'both sides'.

The multivariable survival analyses demonstrated a higher mortality rate for patients undergoing cemented compared with uncemented THR (adjusted HR =1.11, 95% CI 1.07 to 1.16). There was a lower revision rate for cemented procedures (0.53, 0.50 to 0.57). The authors stated that these findings translate into small predicted differences in population averaged absolute survival probability at all time points. At eight years after surgery the predicted probability of death in the cemented group was 0.013 higher (0.007 to 0.019) than the uncemented group and the predicted probability of revision was 0.015 lower (0.012 to 0.017). In multivariable analyses which only included men, there was a higher mortality rate in the cemented group and the uncemented group when compared to the RS group. RS had a

similar revision rate to uncemented THR, hence 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 rather than cemented THR, and a small but significant increased risk of death with cemented procedures.

A study from Smith et al. (2012)[15] reported that in women, RS resulted in worse implant survival than THR regardless of head size. The predicted five year revision rates in 55-year-old women were 8·3% (95% CI 7·2 to 9·7) with a 42 mm RS head, 6·1% (5·3 to 7·0) with a 46 mm RS head, and 1·5% (0·8–2·6) with a 28 mm cemented MoP THR. In men with smaller femoral heads, RS resulted in poor implant survival. Predicted five-year revision rates in 55-year-old men were 4·1% (3·3 to 4·9) with a 46 mm RS head, 2·6% (2·2 to 3·1) with a 54 mm RS head, and 1·9% (1·5 to 2·4) with a 28 mm cemented MoP stemmed THR. Of male RS patients, only 23% (5,085 of 22,076) had head sizes of 54 mm or above. 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.

Finland

Seppanen et al. (2012)[319] analysed the risk of revision of 4,401 RS in the Finnish Arthroplasty Register compared to the risk of revision to 48,409 THRs performed during the same time period. The median follow-up time was 3.5 (0–9) years for RS and 3.9 (0–9) years for THRs. The study reported no statistically significant difference in risk of revision (RoR) between RSs and THRs (RoR = 0.93, 95% CI 0.78 to 1.10). The four-year unadjusted Kaplan-Meier survival was 96% (95% CI 96 to 97) for both RS and THR groups. Female patients had about double the risk of revision compared to male patients (RoR = 2.0, CI 1.4 to 2.7).

Australian

Buergi et al. (2007)[301] reported the use of RS based on the Australian National Joint Replacement Registry. A total of 7,205 RS procedures were implanted between 1999 and 2005. The study concluded that early revision rates for RS were higher in the database than for THR. At three years, the revision rate after RS was 2.8% compared to 2.0% for THR.

Multinational

Corten et al. (2010)[303] compared RS survivorship reported by registries (Australian, England and Wales, Sweden) to failure of THR between 2006 and 2009. RS was associated with an overall increased failure rate in comparison to THR. The results demonstrated the cumulative revision rate reported in the Australian registry (3.7% RS and 2.7% THR). The three-year revision rate for RS was 1.8% in England and Wales and 3.4% for RS in Sweden.

A study using data from the Nordic Arthroplasty Registry compared the outcome of RS (n=1,638) vs. THR (n=309,290) between 1995 and 2007[339]. Results indicated that RS had a three-fold increased revision risk compared to THR (relative risk (RR) = 2.7, 95% CI 1.9 to 3.7). The difference was greater when RS was compared to the cemented THRs (RR = 3.8, CI 2.7 to 5.3). In men less than 50 years of age the difference was less (RS versus THR: RR = 1.9, CI 1.0 to 3.9; RS vs. cemented THR: RR = 2.4, CI 1.1 to 5.3). However it was higher in women of the same age group (RS vs. THR: RR = 4.7, CI 2.6 to 8.5; RS versus cemented THR: RR = 7.4, CI, 3.7 to 15). In the Cox regression analysis, RS showed an increased risk of early aseptic revision compared to THR (RR = 2.7, CI 1.9 to 3.7; $p < 0.001$) and all-cemented THR (RR = 3.8, CI 2.7 to 5.3; $p < 0.001$).

The purpose of one recent study[341] was to evaluate the outcome of the Birmingham Hip Resurfacing (BHR) arthroplasty using revision rates as reported in national joint replacement registry studies (categorised as UK, Australia, Asia and USA). In total 9,806 RS 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 and personal expertise) (0.27; CI 0.14 to 0.40) and the register data (0.74; CI 0.72 to 0.76). The average revision rate from register data was 3.41% (SD 1.79) and corresponded to 0.74 revisions per 100 observed component years (CI 0.72 to 0.76%).

Table 49. Summary table of registry studies for RS

Author	Registry	Implant type/comparator	Outcome report	Result reported
Jameson et al. (2012)[343]	NJR for England and Wales	Men vs. Women 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 et al. (2012)[308]	NJR for England and Wales	Cemented or uncemented procedures versus men undergoing RS	Mortality and revision rates (8 yr)	Higher mortality rate for patients undergoing cemented compared with uncemented THR (adjusted HR =1.11, 95% CI 1.07 to 1.16)
Smith et al. (2012)[15]	NJR for England and Wales	Men versus Women RS by femoral head size	Revision rate (5 yr)	55-year-old women (RR) 8·3% (95% CI 7·2 to 9·7) with a 42 mm RS head, 6·1% (5·3 to 7·0) with a 46 mm RS head, and 1·5% (0·8–2·6) with a 28 mm cemented metal-on-polyethylene THR. 55-year-old men were (RR) 4·1% (3·3 to 4·9) with a 46 mm RS head, 2·6% (2·2 to 3·1) with a 54 mm RS head, and 1·9% (1·5 to 2·4) with a 28 mm cemented MoP stemmed THR
Seppanen et al. (2012)[319]	Finnish Arthroplasty Register	RS vs. THR	Risk of revision (3.5-3.9 yr)	No statistically significant difference in risk of revision (RiR) between RSs and THRs (RR = 0.93, 95% CI 0.78 to 1.10)
Buergi et al. (2007)[301]	Australian National Joint Replacement Registry	RS vs. THR	Risk of revision (3 yr)	Revision rate after RS was 2.8% compared to 2.0% for THR
Corten et al. (2010)[303]	Multinational	RS versus THR	Revision rate (3 yr)	Revision rate for RS was 1.8% in England and Wales and 3.4% for RS in Sweden.
Johnson et al, 2010[339]	Nordic Arthroplasty Registry	RS versus THR	Relative risk	RS had a 3-fold increased revision risk compared to THR (relative risk = 2.7, 95% CI 1.9 to 3.7)
Schuh et AL, 2012[341]	Multinational	RS reported in registry versus 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; CI 0.14 to 0.40) and the register data (0.74; CI 0.72 to 0.76). Average revision rate was 3.41% (SD 1.79). 0.74 revisions per 100 observed component years (CI 0.72 to 0.76%)

RS = resurfacing arthoplasty, THR= total hip replacement, yr= year, RR= revision rate, HR= hazard ration, RiR=risk of revision SD=standard deviation CI=confidence interval

Summary of RS 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 five 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 due to the larger femoral head compared with conventional THR.[341] However, the majority 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.[319] No studies were included that reported RS implant survival as better than all THR. One-study of men only reported that RS had a similar revision rate to uncemented THRs, but that both had a higher revision rate than cemented THRs.[308]

6.9.2.2 Total hip replacement

Evidence on registry studies reporting THR was limited to 22 studies, the majority of these studies investigated various comparison of THR surgery or demographic details regarding the specific country. See Table 50 for a summary of the THR papers.

England and Wales

Jameson et al. (2012)[307] reported survival time to revision following primary cemented THR in 34,721 THRs recorded in the NJR for England and Wales between 2003 to 2010. The authors reported the seven-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, gender, 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. (2012)[16] assessed the use of MoM bearing surfaces in the NJR between 2003 and 2011. They reported that MoM THR failed at high rates and that this was linked to head size. Analysis of the 31,171 MoM THRs larger heads failed earlier (3.2% cumulative incidence of revision 95% CI 2.5 to 4.1, for 28 mm and 5.1% 95% CI 4.2 to 6.2, for 52 mm head at 5 years in men aged 60 years). The five-year revision rates in younger women were 6.1% (95% CI 5.2 to 7.2) for 46 mm MoM compared with 1.6% 95% CI 1.3 to 2.1, for 28 mm MoP. This finding contrasted with CoC bearing surface where larger head sizes were associated with improved survival (5 year revision rate of 3.3% 95% CI 2.6 to 4.1, with 28 mm and 2.0% 95% CI 1.5 to 2.7, with 40 mm for men aged 60 years).

Denmark

Johnsen et al. (2006)[329] examined the association between patient-related factors and the risk of initial, short and long-term failure after primary THR using data from the Danish Hip Arthroplasty Registry (n=36,984). The study concluded that in Denmark male gender and co-morbidity index score (Charlson index) were strongly predictive of THR failure between 1995-2002. The Charlson index includes 19 disease categories which correspond to ICD-8 and ICD-10 codes used in the national registries. A total of 1,132 primary THRs were revised (3.1% of the 36,984 procedures) during this time period.

A more recent study from Denmark evaluated short-term (0 to 90 days) and longer term, (up to 12.7 years) mortality of patients undergoing primary THR compared to the general population[331]. THR (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 one-month period of increased mortality immediately after surgery among THR patients (adjusted mortality rate ration 1.4 (95% CI 1.2 to 1.7), however, overall short-term mortality (0 to 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 under 60 years old, and among THR patients without comorbidity. Long-term mortality was lower among THR patients than in the general population controls (adjusted mortality rate ratio 0.7; 95% CI 0.7 to 0.7).

Sweden

Lazarinis et al. (2010)[333] analysed patient data (n=8,043) on cementless cups with or without hydroxyapatite (HA) coating that had been recorded in the Swedish Hip Arthroplasty Register between 1992 and 2007. The primary endpoint was revision due to aseptic loosening, the secondary endpoints were cup revision for any reason, and cup revision due to infection. The results reported that HA coating was a risk factor for cup revision due to aseptic loosening (adjusted RR 1.7; 95% CI 1.3 to 2). Age at primary THR of < 50 years, pediatric hip disease, cemented stem, and the cup brand were also associated with statistically significantly increased risk of cup revision due to aseptic loosening.

A more recent study from Sweden reported data from 1999 to 2010.[334] The authors investigated revision rates of monoblock cups used in primary THR that were registered in the Swedish Hip Arthroplasty Register. 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 five-year survival with any revision as the endpoint was 95% (95% CI 91 to 98) for monoblock cups and 97% (CI 96 to 98) for modular cups (p = 0.6). The adjusted HR for revision of monoblock cups compared to

modular cups was 2 (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. (2012)[304] 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 $4.1 billion. Results state that if the poor performing THRs had been conducted using average longevity designs, the number of THR revisions could have reduced by 47%.

One study investigated the relationship between the bearing surface and the risk of revision due to dislocation using 110,239 records of the AOANJRR between 1999 and 2007.[305] They reported that 2,621 (2.4%) of all primary THRs, were revised for any reason; 862 (0.78%) THRs were revised because of dislocation. CoC bearing surfaces had a lower risk revision due to dislocation than MoP and CoP at seven years follow-up. They reported a significantly higher rate of revision for dislocation in CoC bearing surfaces than in MoP 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) and in older patients (≥ 65 years) (HR = 1.73, p = 0.016).

Italy

DiTanna et al. (2011)[260] report from 2000 to 2007 from the Emilia-Romagna Regional Registry on Orthopaedic Prosthesis (RIPO), which collects information on all orthopaedic intervention performed in Emilia-Romagna Italy. The study assessed the cost-effectiveness of cementless versus hybrid prostheses in 41,199 THRs and concluded differences in the revision rates and impact upon cost. They concluded that, considering two cohorts of 100 subjects, 243 revisions would be expected in the cementless group versus 300 in the hybrid group. This was equal to a 19% difference and a number needed to treat (NNT) of 18.

A second paper reporting on the RIPO[310] conducted survival analysis using the Kaplan Meier method to analyse survival rates of THR in Italy between 2000 and 2006 (35,042 THRs, 5,878 revisions). The reported cumulative survival rate at seven years was 96.8% (95% CI 96.4 to 97.1) for THR. Multivariate analysis demonstrated that THR was affected by pathology, e.g. the presence of rheumatoid arthritis. Women comprised 66.4% of patients and more than 54.0% were overweight (BMI higher than 25). Mean age at primary surgery was 66.9 years (range, 16 to 101 years), at revision 70.0 years (22 to 98).

Finland

Eskelinen et al. (2005)[313] evaluated the population-based survival of cementless THR in patients under 55 years of age using data from the Finnish Arthroplasty Register. All cementless stems studied showed a survival rate of over 90% at 10 years.

Makela et al. (2001)[317] analysed population-based survival rates for cemented and cementless THRs in patients aged 55 years or over in Finland between 1980-2006. The 15 year survival rate of cementless THR (80%) was comparable with the rates of 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 cups, analysed separately, had a significantly lower risk of revision for aseptic loosening than did cemented implants.

The same authors reported revision outcome for primary OA.[318] The 15-year survival for any reason of cementless THR group one (implants with a cementless, straight, proximally circumferentially porous-coated stem and a porous-coated press-fit cup) operated on 1987–1996 (62%; 95% CI 57 to 67) and cementless group two (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) (58%; CI 52 to 66) operated on during the same time period was worse than that of cemented THRs (71%; CI 62 to 80), although the difference was not statistically significant. The risk of revision for aseptic loosening of cementless stem group one operated on 1987–1996 (0.49; CI 0.32 to 0.74) was lower than that for aseptic loosening of cemented stems ($p = 0.001$).

Slovakia

One study reported findings from Slovakia[335] from 2003 to 2010, a total of 4,970 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 hybrid. By 2010, the revision rate reached 9.20%, representing an annual increase of 1.1%. The revision rate in the whole observed period 2003 to 2010 was 9.15%.

Norway

Espehaug et al. (2011)[322] studied differences by county and regional health authority over a 20 year period (1989 to 2008) using data from the Norwegian Arthroplasty Register. The authors observed an increase in THR from 109 operations per 100,000 inhabitants in the years

1991–1995 to 140 in 2006–2008. Variations were found across the four regions studied in Norway.

A second study from Norway[323] reported the results after THR in terms of revision rate, during a 21-year period among hip replacements reported to the Norwegian Arthroplasty Register. Risk of revision during the time periods 1993–1997, 1998–2002, and 2003–2007 was compared to that of the reference period 1987–1992. There was an overall reduced risk of revision in the time periods 1993–1997, 1998–2002, and 2003–2007 compared to the reference period. The improved results were due to a reduction in 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 done 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.

A study to compare the difference in the risk of revision for infection and changes in risk over time and in time from primary surgery to revision for infection after THR was conducted using data from the Norwegian Arthroplasty Register (1987 to 2008).[325] The study report the risk of revision for infection from six years postoperatively was higher in patients with RA. Of the 84,492 THRs, 534 (0.6%) were revised for infection. Women had a significantly lower risk of revision for infection compared with men (RR 0.41, 95% CI 95% 0.34 to 0.48). The cumulative five-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 versus OA patients) with revision for infection as the end point.

USA

One study reported registry data from the USA.[327] 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 to 2009. A total of 30,491 THRs were included in the analysis, of these 17,474 (57%) were performed on women. The incidence of SSI was 0.51% (155 of 30,491). A total of 155 deep SSIs (0.51% (95% CI 0.43% to 0.59%)) occurred at a mean of 72 days (median 28 (SD 93.3)) after the procedure. Patient factors associated with SSI included female gender, obesity, and (ASA) grade ≥ 3.

Multinational

Sadoghi et al. (2012)[340] compared primary THRs between different countries in terms of THR number per inhabitant, age, and procedure type and to compare the survival curve including

all THRs using hip arthroplasty registers. The analysis used data from nine registers. On average the annual number of primary THRs per 100,000 inhabitants was found to be 133 for all ages, 26 for persons younger than 55 years, 269 for persons 55–64 years, 520 for persons 65–74 years and 531 for persons older than or equal to 75 years. The fixation method varied by country, e.g. in Sweden 67% are cemented THRs whereas in Emilia-Romagna (Italy) 89 % are cementless THAs. Cementless fixation was more popular in Australia, Denmark, Emilia-Romagna, New Zealand, and Portugal (50%). Cemented fixation was used more in Sweden and Norway (50%). Cemented and cementless fixations were used equally in England and Wales and Slovakia. Hybrid fixation was more equal across countries, and ranged between 8 % in Portugal to 34.5% in New Zealand. Denmark showed the lowest survival rate within the first 15 years; however, THRs used between 2006 to 2009 in Norway had similar survival rates. All survival curves calculated in the study (except for Denmark) varied less than 1% within nine years. Multivariate or subgroup analyses were not performed to compare the survival curves. The use of primary RS was not separated in the registries from Norway and Slovakia. Use of RS varied between 1% in Portugal, 2 % and 3% in Denmark, Emilia-Romagna, New Zealand and Sweden to approximately 5% and 6% in Australia and England and Wales.

Graves et al. (2011)[337] performed an investigation of the use of MoM THR 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 sizes when MoM bearing surfaces.

The Nordic registry included the joint registries of Denmark, Sweden, and Norway. One study [338]aimed to compare demographics, choice of implant, fixation techniques, and results between the countries, a total of 280,201 THR between 1995-2006. Results reported 9,596 (3.4%) of the THRs had later been revised. RS of hips accounted for 0.5% or less in all countries. Ten-year survival was 92% (95% CI: 91.6–92.4) in Denmark, 94% (95% CI: 93.6–94.1) in Sweden, and 93% (95% CI: 92.3–93.0) in Norway.

A second study reporting data from the Nordic registry compared the survival of cemented THR with metal femoral heads made from various materials (cobalt-chromium, aluminium and zirconium).[324] The study reported the prosthesis survival and relative revision risks adjusting for age, sex, and diagnosis between 1987 and 2010. There were 132,000 cases of THR included in the analysis. At 12 years, the survival rate was 88.1% with cobalt-chromium heads and 74.8% with zirconium heads. Aluminium femoral heads provided no advantage over cobalt-chromium heads on prosthesis survival. They concluded that cemented

polyethylene THR with aluminium heads had similar survival as the same THR with ceramic-on-ceramic heads when any revision was the end point.

Table 50. Summary table of registry studies for THR

Author	Registry	Implant type/comparator	Outcome	Reported result
Jameson et al. (2012)[307]	NJR for England and Wales	Primary cemented THR	Survival time to revision (7yr)	7 year rate of revision for any reason as 1.70%
Smith et al. (2012)[16]	NJR for England and Wales	MoM vs. THR – head size and gender	Survival time to revision (5yr)	Larger heads failed earlier 3.2% cumulative incidence of revision 95% CI 2.5 to 4.1, for 28 mm and 5.1% 95% CI 4.2 to 6.2, for 52 mm head at 5 years in men aged 60 years. The five year revision rates in younger women were 6.1% (95% CI 5.2 to 7.2) for 46 mm MoM compared with 1.6% 95% CI 1.3 to 2.1, for 28 mm MoP
Johnsen et al. (2006)[329]	Danish Hip Arthroplasty Registry	Patient-related factors and the risk of initial, short and long-term failure after primary THR	Implant revision	Male gender and co-morbidity index score (Charlson index) were strongly predictive of THR failure. 3.1% of the 36,984 procedures were revised
Pedersen et al. (2011)[331]	Danish Hip Arthroplasty Registry	Mortality of patients undergoing primary THR compared to the general population	Adjusted mortality rate ratio	Long-term mortality was lower among THR patients than in the general population controls (adjusted mortality rate ratio 0.7; 95% CI 0.7 to 0.7)
Lazarinis et al. (2010)[333]	Swedish Hip Arthroplasty Register	Cementless cups with or without hydroxyapatite (HA)	Revision due to aseptic loosening	HA coating was a risk factor for cup revision due to aseptic loosening (adjusted RR 1.7; 95% CI 1.3 to 2)
Weiss et al. (2012)[334]	Swedish Hip Arthroplasty Register	Monoblock cups vs. modular cups	Implant survival (5yr)	95% (95% CI 91 to 98) for monoblock cups and 97% (CI 96 to 98) for modular cups (p = 0.6)
Luo et al. (2012)[304]	Australian Orthopaedic Association National Joint Replacement Registry	Identification of implants with higher than expected failure rates between 2003 and 2007	N/A	Results state that if the poor performing THRs had been conducted using average longevity designs, the number of THR revisions could have reduced by 47%
Sexton et al. (2009)[305]	Australian Orthopaedic Association National Joint Replacement Registry	MoP vs. CoC	Rate of revision	Higher rate of revision for dislocation in CoC, than in MoP 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) and in older patients (≥ 65 years) (HR = 1.73, p = 0.016)
DiTanna	Emilia-	Cementless vs.	Number of	243 revisions would be expected in

et al. (2011)[260]	Romagna Regional Registry on Orthopaedic Prosthesis	hybrid prostheses	revisions expected	the cementless group vs. 300 in the hybrid group. This was equal to a 19% difference and a NNT of 18
Stea et al. (2009)[310]	Emilia-Romagna Regional Registry on Orthopaedic Prosthesis	Analyse survival rates of THR in Italy between 2000 and 2006	Implant survival rate (7yr)	seven year implant survival was 96.8% (95% CI 96.4 to 97.1)
Eskelinen et al. (2005)[313]	Finnish Arthroplasty Register	Population-based survival of cementless THR	Implant survival rate (10yr)	Survival rate of over 90% at 10 years for cementless THR
Makela et al. (2001)[317]	Finnish Arthroplasty Register	Cemented vs. cementless THR	Implant survival rate (15yr)	15 year survival rate of cementless THR (80%) was comparable with the rates of the cemented groups (86%)
Makela et al. (2001)[318]	Finnish Arthroplasty Register	Cemented vs. cementless THR for OA patients	Implant survival rate (15yr)	Cementless THR were worse during the same time period (62%; 95% CI 57 to 67) and (58%; CI 52 to 66) than cemented THRs (71%; CI 62 to 80)
Necas et al, (2011)[335]	Slovakia	Operation performed between 2003 to 2010	Revision rate (7yr)	revision rate in the whole observed period 2003 to 2010 was 9.15%
Espehaug et al. (2011)[322]	Norwegian Arthroplasty Register	Differences by county and regional health authority over a 20 year period (1989 to 2008)	Numbers of THR performed	Increase in THR from 109 operations per 100,000 inhabitants in the years 1991–1995 to 140 in 2006–2008
Fevang et al, 2010[323]	Norwegian Arthroplasty Register	Risk of revision during the time periods 1993–1997, 1998–2002, and 2003–2007 was compared to that of the reference period 1987–1992	Revision risk	Reduced risk of revision in the time periods 1993–1997, 1998–2002, and 2003–2007 compared to the reference period
Schrama et al, 2010[325]	Norwegian Arthroplasty Register	THR in RA patients vs. OA patients	Implant survival (5yr)	Five-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 et al. (2012)[327]	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)) after the procedure
Sadoghi et al. (2012)[340]	Multinational	Compared primary THRs between different countries in terms of THR number per inhabitant, age, and procedure	Implant survival	Denmark showed the lowest survival rate within the first 15 years. Norway had similar survival rates

		type		
Graves et al. (2011)[337]	Multinational	The use of MoM THR across three registries	N/A	All registries reported an increased revision rate associated with larger femoral head sizes when MoM bearing surfaces
Havelin et al. (2009)[338]	The Nordic Registry	Compare demographics, choice of implant, fixation techniques, and results between countries	Implant survival (10yr)	10-year survival was 92% (95% CI: 91.6–92.4) in Denmark, 94% (95% CI: 93.6–94.1) in Sweden, and 93% (95% CI: 92.3–93.0) in Norway
Kadar et al. (2012)[324]	The Nordic Registry	Metal femoral heads made from various materials (cobalt-chromium, aluminium, zirconium)	Implant survival (12yr)	The survival rate was 88.1% with cobalt-chromium heads and 74.8% with zirconium heads

NNT=number needed to treat RS = resurfacing arthoplasty, THR= total hip replacement, yr= year, RR= revision rate, HR= hazard ration, RiR=risk of revision SD=standard deviation CI=confidence interval, SSI=surgical site infection

Summary of THR studies

The 22 THR studies reported registry data analysed across nine countries. These studies examined various aspects of the THR procedure, including revision and survival rates, comparison of varying implants and combinations of implants, outcome measures such as reason for failure and patient differences associated to failure. Four of the 22 THR studies used registry data from multinational databases. Sadoghi et al. (2012)[340] 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 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 Denmark showed the lowest survival rate within the first 15 years.

6.9.2.3 Core articles included in the economic model and survival analysis

The prioritisation of the eligible studies identified 30 papers that were determined as potentially useful for the economic model and survival analysis. The final number of core papers which help to inform the discussion of survival analysis in this report was three. [15,16] [308] This was in addition to the annual report from the Swedish Arthoplasty Registry, the NJR 2011 report and the AOANJRR which were used for comparison of survival analysis methods.[94] [36] [93]

6.10 Summary of overall registry evidence

The review of registry studies resulted in 30 core papers that were included for essential information. 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 was 22, they reported various aspects of the THR procedure, including revision and survival rates, however these varied between three and 15 years. Comparison of varying implants and combinations of implant bearing surfaces was conducted. Additional outcome measures included reason for failure (e.g., infection) and patient/demographic differences associated to failure.

7 INDIVIDUAL PATIENT DATASET

7.1 Introduction to IPD analysis

This chapter provides a narrative description of the individual patient data (IPD) that was retrieved from the NJR and used for analysis in this report. The dataset is known here as the NJR, and data comes from the 009 data set with primary operations done before 1/3/2012. Any revision or notified death up to September 2012 was been included. The NJR is maintained on behalf of Department of Health and Welsh government, it was established in 2002 and is updated annually, hip and knee joint replacements were collected from April 2003. Northern Ireland joined in 2013, this 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 reports on data from some of the private operations carried out in independent hospitals.

7.2 Method

This is a retrospective cohort study, which involves analysis of NJR data in order to derive time to revision of hip replacement procedures. The data provided by the NJR was divided into two types, depending on type of surgery i.e., RS and THR. THR data was separated into five categories on the basis of the frequency of combinations of the components used in the procedure.

7.3 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 in order to evaluate cost-effectiveness of the THR and RS economic models (see Chapters 9 & 10). For the purpose of this report and in line with the scope, information and analyses have been stratified by procedure type (THR and RS).

7.4 Structure of the database

The NJR database collects numerous variables relating to patients joint affected, outcomes, procedures and implants. For the purposes of this study 198 variables were requested from both the RS and THR databases. The extracted data contained the following information:

1. Patient demographics
2. Provider type
3. Lead surgeon grade
4. Procedure types/patient procedure/side
5. Indications for primary surgery

6. Primary thromboprophylaxis
7. Primary untoward intra operative events
8. Primary bone graft usage
9. All primary implant details
10. Current outcome type
11. Time from primary operation to outcome
12. Age at death
13. 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 rheumatoid arthritis sero-negative or rheumatoid arthritis sero-positive. These were excluded from analysis of time to revision.

7.5 Contents of the database

In order to evaluate cost-effectiveness of hip replacement procedures in line with the scope, we requested the following variables for the two patient groups (RS and THR) separately:

- RS - which 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)
- THR - which involves the removal of the entire damaged hip joint and replacement with an artificial joint (n=387,694)

7.6 Results

The primary outcome was time to revision for statistical and economic modelling.

7.6.1 Hip RS arthroplasty

This section describes the data reported for the patients in the NJR RS dataset. Figure 20 shows outcomes for this group of patients. Of 31,222 patients 9,339 were female and 21,883 male. Further subdivision according to age and head size is shown in s

Table 51 and Table 52.

Figure 20. Endpoint for all RS included in the analysis

Table 51. Total number of male gender by head size for RS

MALES	Head Size													
Age Group	36	38	40	42	44	46	48	50	52	54	56	58	60	Total
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	2,113
45-54	0	2	3	13	89	565	936	2,516	1,109	1,312	164	121	3	6,833
55-64	1	1	5	22	123	776	1,334	3,717	1,519	1,882	204	150	4	9,738
65-74	0	0	1	9	24	206	340	1,070	404	564	87	47	3	2,755
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	1,812	2,969	8,223	3,396	4,251	505	358	11	21,883

Table 52. Total number of female gender by head size for RS

FEMALES	Head Size												
Age Group	34	36	38	40	42	44	46	48	50	52	54	58	Total
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	1,267	240	225	22	14	1	3,350
55-64	0	1	31	133	738	759	1,678	355	342	20	9	1	4,067
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	1,928	1,689	3,683	748	714	55	27	2	9,339

7.6.2 Total hip replacement

The NJR describes the results of the patients undergoing THR surgery in England and Wales from April 2003 and December 2012. On date of receipt of the data (06.12.2012) the dataset had a total of 387,694 records. From this number only 387,667 records were usable due to following reasons:

1. Irrelevant data type reported (negative age, zero age) (22 records)
2. Missing variable information (11 records)

The remaining 387,667 patients could have one of three outcomes: (Figure 21)

1. Death
2. Unrevised THR
3. Revision surgery

Figure 21. End point of all THR included in the analysis

Of these 387,667 patients 240,156 (62%) were selected for analysis on the basis of frequency of use of different THR components and from these 239,089 patients with 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.

7.7 THR category development

The NJR database for non-RS contained 387,691 records. After removing 24 records (this included records with missing entries and where the primary time to outcome was negative), the database contained 387,667 useable records.

The database contained several key components for hip replacement which was used to determine the categories which 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 tabs for each of the variables listed above to determine the most frequent combinations. For example, for the cup component group we cross tabbed this with the liner composition. These two-way cross tabs were done for all the covariates listed above.

We then added another component which was the most frequently occurring. For example, looking at the two-way cross tab for cup component, group and head composition, we know from the previous two-way cross tab 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 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 more than 25,000 operations. Our expert clinical advisor identified a further exclusive category (n = 12,773) which is a well-known option consisting of a cemented stem with a ceramic head articulating with a cemented polyethylene cup. 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 (See Figure 22).

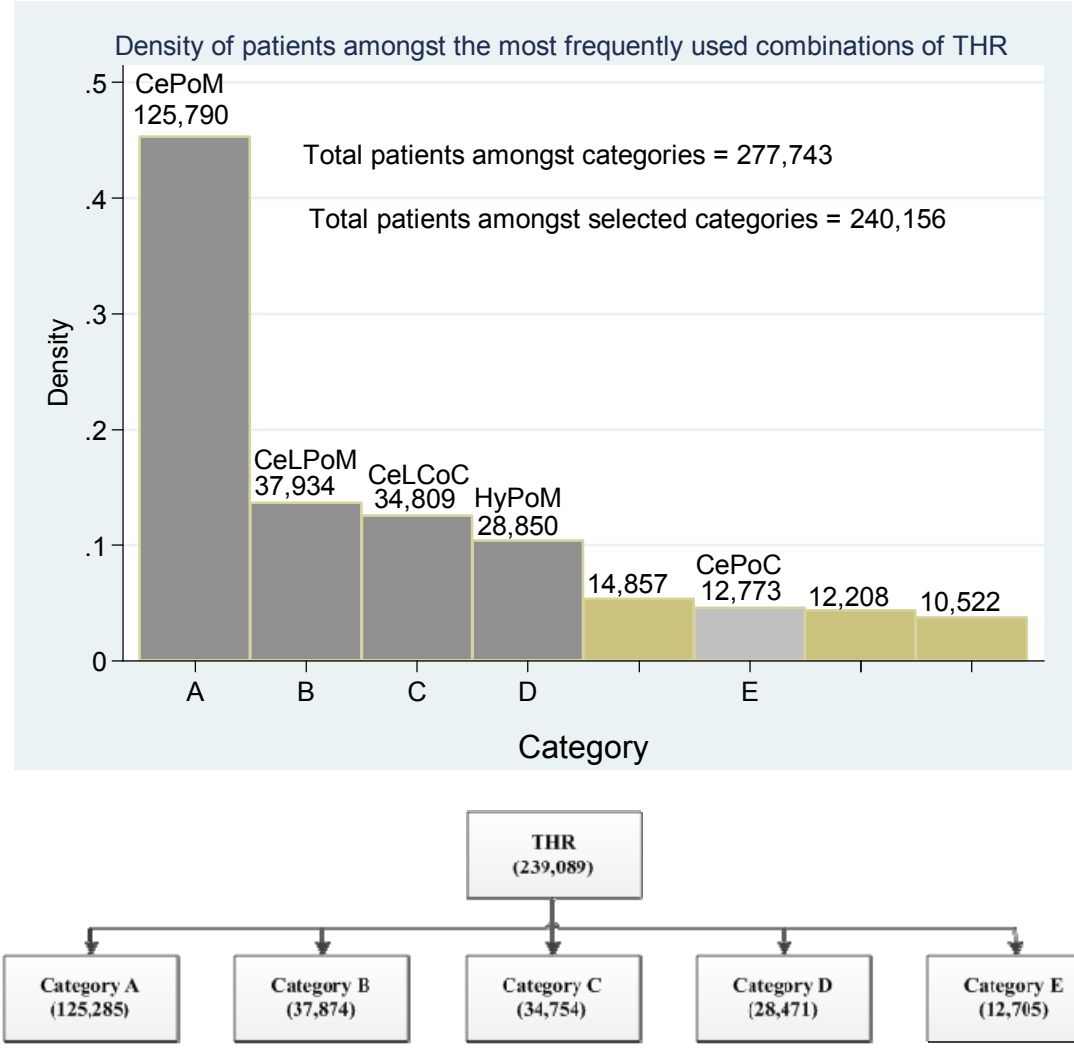


Figure 22. The frequency of each of the five categories of the THR dataset

The table below shows the final five categories which we have used in our 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 below with their associated short form acronyms (

Table 53). Further information on age and gender distribution and technical characteristics of the categories is provided in Table 54 and Table 55.

Table 53. Characteristics of the five THR categories

Categories	Characteristics	Acronym for use in the report
A	Cemented poly cup on metal head (cemented stem)	CePoM
B	Cementless HA coated metal cup (poly liner) on metal head (cementless stem)	CeLPoM
C	Cementless HA coated metal cup (ceramic liner) on ceramic head (cementless stem)	CeLCoC
D	Cementless HA coated metal cup (poly liner) on metal head (cemented stem)	HyPoM
E	Cemented polyethylene cup on ceramic head (cemented stem)	CePoC

Ce= cemented CeL=cementless Hy=Hybrid P=polyethylene M=metal C=ceramic

Table 54. Constitution of THR categories by age and gender

	Female over 65 years old	Male over 65 years old	Females less than 65 years old	Males less than 65 years old	Total
A	75,734	37,018	8,079	4,454	125,285
B	18,396	11,878	4,423	3,177	37,874
C	7,554	6,186	11,698	9,316	34,754
D	15,641	8,657	2,649	1,524	28,471
E	4,655	2,777	3,073	2,200	12,705
Total	121,980	66,516	29,922	20,671	239,089

Table 55. Prosthesis characteristics for the five categories of THR

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	Count with only OA patients
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

7.8 Matching

In health evaluation, data often do not come from randomized trials but from (non-randomised) observational studies. Rosenbaum and Rubin (1983) proposed propensity score matching as a method to reduce the bias in the estimation of treatment effects with observational data sets.[344] Propensity matching on age and gender was undertaken using the *Edwin Leuven* procedure.[345]

The rationale for using propensity scores is that since 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 the way to adjust or correct the estimation of treatment effects, controlling as far as possible for the existence of confounding factors and 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 1/3/2012.

We combined data for men and women of all ages for RS (31,222 patients) and for THR (239,594 patients). We selected patients aged less than 65 years for matching; 9,339 females and 21,883 males from the RS group were matched with those from the THR group.

Analysis to match the RS and THR groups was performed using the Statistical package Stata 12 Special Edition (StataCorp LP, StataCorp 4905 , Lakeway Drive College Station, Texas 77845 USA 800-STATA-PC).

We used the Stata command "psmatch2".[345] We used Nearest-Neighbor Matching (NN) using one-to-one matching by identifying the 'nearest neighbour' to each RS patient from the THR database based on closest propensity score; variables used to construct the propensity score were age and gender.

In using these programs, it should be kept in mind that they only allow us to reduce, and not to eliminate, the bias generated by unobservable confounding factors.

7.9 Assessment of utility and quality of the NJR database

This section considers the utility and quality of the dataset from the perspective of the requirements for the present report. Unsurprisingly, the database structure of this resource was not tailored specifically for the task in hand. The strengths and weakness of the datasets are briefly summarised below:

Strengths:

1. The database was comprehensive in that it contained information on all patients listed for hip arthoplasty 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 THR dataset)
3. The size of the dataset was large, this provides narrow confidence intervals for survival analysis and hence more certainty in the evaluation of the cost effectiveness
4. It was possible to distinguish between THR and SR patients

Weaknesses:

1. The elapsed time to any primary outcome was reported in years rather than number of days or dates
2. There were no costs reported for the procedures
3. It was not possible to link patients that proceeding from the RS to THR dataset
4. Our dataset was not linked by revision surgery
5. There was very poor reporting of BMI
6. There was no linkage to the PROMs dataset in our data

7.10 Summary of individual patient dataset

The NJR provides valuable information about patient sub-groups and the categorisation of hip replacement procedures for all patients receiving treatment in the NHS in England and Wales. There was insufficiently complete data to estimate linked primary and secondary surgery for each patient or costs or utilities associated with the procedures.

Subsequent sections describe further analysis of this database in the cost-effectiveness model.

8 PATIENT REPORTED OUTCOME MEASURES

8.1 Quality of life and utilities

8.1.1 Background

This section provides a brief description of the patient reported outcome measures dataset which was used to provide utility data for analysis in the Markov model (see Chapter 10). We obtained quality of life data from the database of PROMS for patients who had a THR between January 2009 and December 2012 (PROMS, NHS Information Centre, 21st March 2013). The variables in the dataset included the following: PROMS ID, patient gender, patient death, surgery date, complications (e.g. bleeding, infection and wound problems, readmission, further surgery) and EQ-5D-3L data which was completed six months after surgery.

The EQ-5D-3L is a generic health-related quality of life which 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, severe problems. This creates 243 possible health states, to which unconscious and dead have been added for a total of 245 in all. These health states are then converted to an index score from 0 (dead) to 1 (perfect health) value scale.

8.1.2 Methods

Two analyses were undertaken of the PROMS dataset:

Analysis 1

The PROMS dataset for patients who had a THR between January 2009 and December 2012 included 207,436 records. After removing missing EQ-5D scores or surgery dates the dataset contained 117,044 records. No age-specific utilities by gender were available in this dataset.

Analysis 2

A second PROMS dataset containing EQ-5D-3L data for THR by age and gender for the year 2010/2011 was downloaded from the information centre website in March 2013 (http://www.ic.nhs.uk/catalogue/PUB07049) for further analysis. This dataset included 38,378 records. After removing patients with missing information on EQ-5D scores, gender and age category, and after excluding patients under the age of 40 years, the dataset contained 32,577 records.

Overall

For both analyses, mean EQ-5D index results including SD and 95% CIs were calculated. All statistical analyses were conducted in Stata version 12.[346]

8.1.3 Results

Table 15. EQ-5D utility index scores for all patients who completed the EQ-5D-3L questionnaire after total hip replacement (Analysis 1)

	All patients	Male	Female
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

For all patients, the mean EQ-5D score after their hip operation was 0.767 (see Table 15). Men had a slightly higher EQ-5D utility index score than women (0.787 vs. 0.753).

Table 16 EQ-5D utility index scores for patients who completed the EQ-5D-3L questionnaire after total hip replacement who required further surgery (Analysis 1)

	All patients	Male	Female
N	3,096	1,320	1,776
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 16 shows that the mean EQ-5D score for patients who required further surgery after hip replacement was 0.575 for men and 0.553 for women.

Table 17. EQ-5D results for all patients by age band and gender who completed the EQ-5D-3L questionnaire after total hip replacement (Analysis 2)

	All patients	**Male**	**Female**
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	4,352	1,883	2,469
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	4,758	6,348
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	4,841	7,467
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	4,017	1,234	2,783
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

Table 17 shows EQ-5D results for patients after surgery for the period 2010/2011 split by gender 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 to 70 years gave a slightly higher value to their health-related quality of life than for any other age band; likewise, women in the age band 60 to 70 years gave a slightly higher value to their health related quality of life than any other age band.

8.1.4 Summary of PROMs

The PROMS dataset has provided valuable information on EQ-5D data by age and gender for use in the economic model for patients who have undergone a THR. However, there was insufficient linkage data to link the PROMS dataset to the NJR dataset.

9 MODELLING REVISION RATES

9.1 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 gender 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 7) and patient cohorts were analysed for time to revision. Kaplan-Meier (KM) and competing risk analysis (CR), were implemented in Stata version 11. For KM analysis, non-revision by end of follow-up and death were censored, for CR analysis the competing risk was death and the risk of interest was revision according to the Stata user-written routine.[347]

KM analyses were fitted with parametric distributions to allow for extrapolation beyond the observed data. Following the NICE Decision Support Unit (DSU) recommendation, the IPD was fitted with Weibull, Gompertz, loglogistic, lognormal 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. Since decreasing hazard is unlikely to capture increasing likelihood of revision due to wear and tear, particularly for those active or of young age, further alternative models (bath tub, Rayleigh and Mitscherlitch) were explored to allow for increasing hazard of revision beyond the observed data. An initial analysis of these was done using ordinary least squares in Stata or Excel. The Rayleigh model predicts a linearly increasing hazard, the bath tub a U shaped hazard, and Mitscherlich a hazard that increases at a decreasing rate with time to reach an asymptote.[348,349]

Rayleigh: $$h = a + 2bt$$

Bath tub: $$h = at + \frac{b}{(1+gt)}$$

Mitscherlich: $$h = \pi - b \exp(lt)$$

(π, *a, b, g* and *l* are constant parameters, *t* is time)

In practice the Mitscherlich and Rayleigh models generated poor fits and were not pursued. The results from Weibull, Gompertz, loglogistic, lognormal and bath tub models for each cohort are

catalogued in Appendix 18 which presents modelled time to revision and hazard for the observation period and for extrapolation to 50 years.

The selection of appropriate model or models for use in the economic analysis was based on Akaike's information criteria (AIC), judgement of plausibility of resulting extrapolations, visual goodness of fit to the IPD-derived KM plot, and plots of log K-M estimated cumulative hazard vs. log modelled cumulative hazard.[350] In gender stratified sensitivity analyses parametric fits were adjusted for age with age for each cohort was centred near the mean. The bath tub models were analysed using the Stata *stgenre* package developed by Crowther and Lambert.[351] This provided considerable advantages including the use of IPD, adjustment for age, prediction of hazard and survival, generation of AIC estimates for comparison with other models and of 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.[352,353]

9.2 Revision rates

9.2.1 Categories of THR

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 main features of these five categories are detailed in

Table 53 Chapter 7.

9.2.2 Patient populations to be compared

The remit from NICE for this report specified the following comparisons in people with pain and disability resulting from arthritis of the hip for which non-surgical management has failed:

a) Different types of primary THR compared with RS for people in whom both procedures are suitable

b) 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 7. The five categories account for ~62% of all NJR THR recipients.

We used NJR data to investigate revision rates.
Figure 23 shows age distribution, according to decade, of NJR patients who received THR or RS respectively, and age distribution by gender for those undergoing RS.

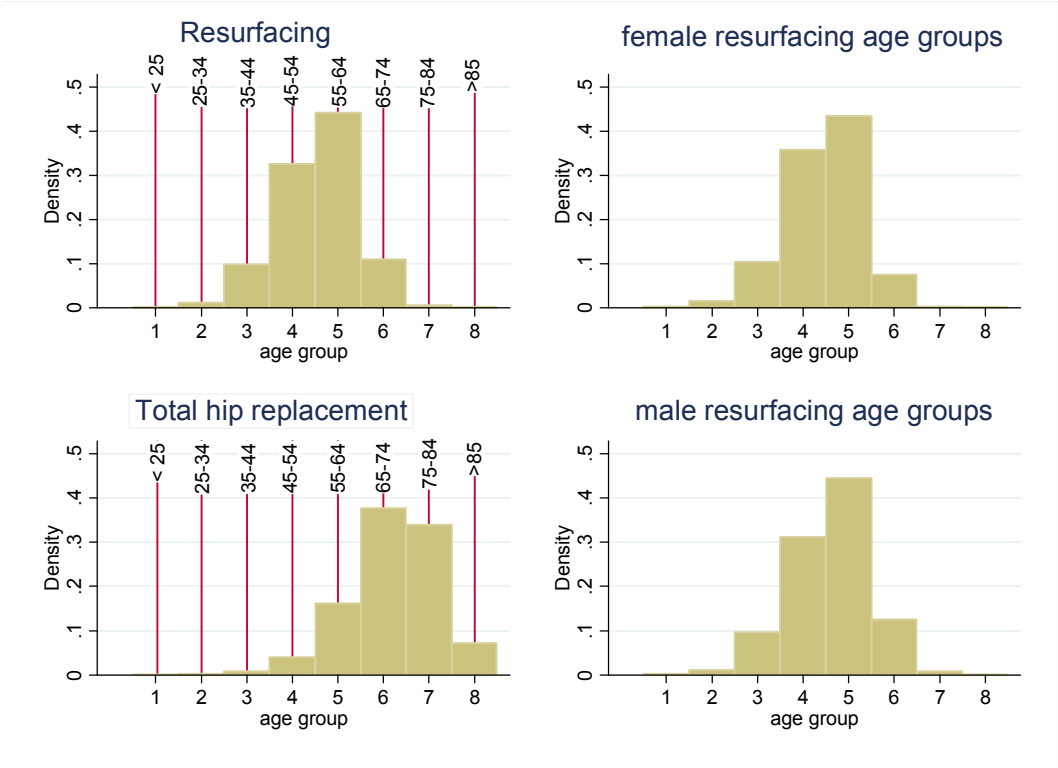


Figure 23. Age distributions of NJR patients receiving THR or RS, and by gender for RS

Most RS patients were aged < 65 years at time of intervention, whereas most THR recipients were older than 65 years. Figure 24 is a Kernel density diagram showing the overlap between the two distributions. We found that populations undergoing RS and THR overlapped substantially (for RS 89.7% were less than 65 years old and for all THR categories 22.6% were less than 65 years old).

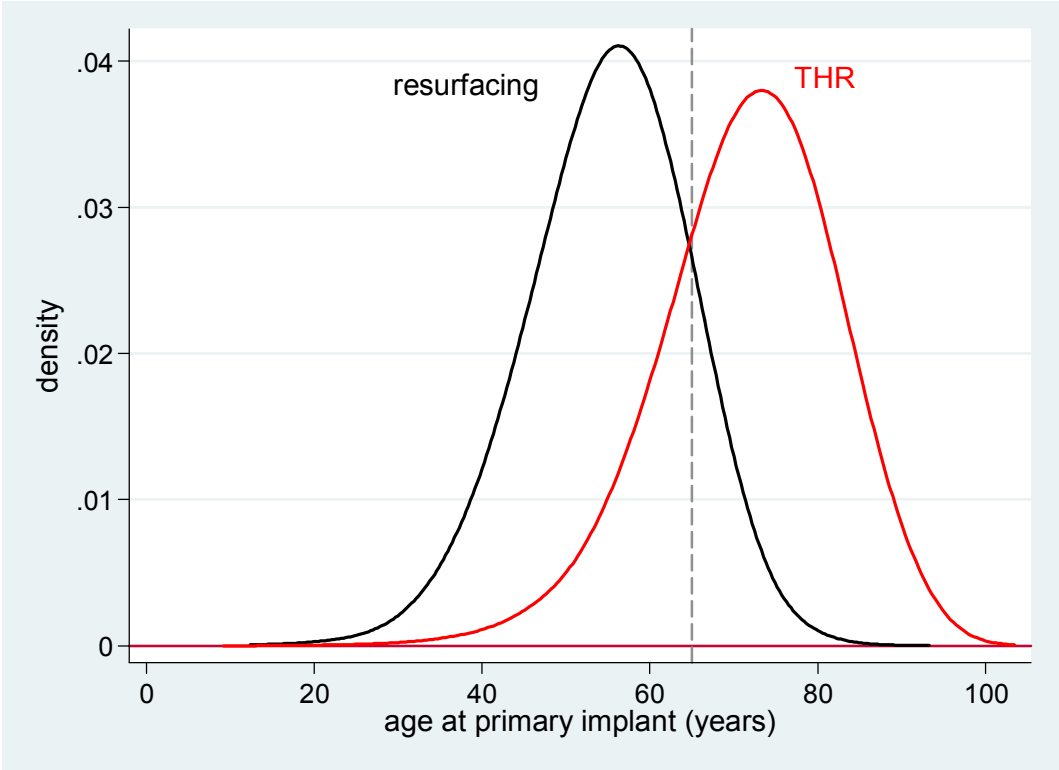


Figure 24. Kernel density diagram of the two distributions

Table 56 summarises the age and gender differences between the populations that received RS and THR. THR interventions outnumbered RS by more than 10:1, the proportion of females was twice as large for THR, and the mean age for RS recipients was about 15 years less than that for THR.

Table 56. Age and gender of RS and THR recipients

Population	Number	% Female	Mean age (SD)	Median	Inter Quartile range
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 Categories 1 to 5 recipients	239,089	63.5	71.6 (9.6)	72.5	65.8-78.3

In order 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 while

THR was the predominant option for less active older patients. The NJR did not provide information on activity levels of patients.

The literature indicates that revision rates after RS are much higher for females than males [15], whereas for THR the reverse is the case, a finding we confirmed in our preliminary analysis (Appendix 19). It is known that revision rates in general are lower for older patients. Because revision rates differ by gender and age it is likely that the cost effectiveness of interventions will reflect the age and gender mix of the population(s) examined. Given the observed differences in age and gender for RS and THR populations the following alternative strategies were considered in order to identify appropriate RS and THR populations for comparison of the interventions:

i] All RS recipients versus all THR recipients not matched
ii] All RS recipients versus all recipients of the 5 identified THR categories not matched (see Chapter 7)
iii] All RS recipients versus each of the (different 16+) categories of THR in the NJR data seta separately matched by age and gender
iv] All RS recipients versus THR recipients from each of the 5 identified categories separately matched by age and gender
v] All RS recipients versus all THR recipients from the combined 5 identified categories matched by age and gender
vi] All RS recipients versus the total pool of all THR recipients matched by age and gender

Options i] and ii] (without matching) were rejected because of the large age and gender 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 iii] to vi] avoid age and gender mismatch if age matching is undertaken separately for each gender and then the matched male and female populations combined. Age matching within genders was in general feasible because of the much larger number of THR recipients than RS recipients. Therefore, we judged options iii] to vi] to be preferable to options i] and ii].

Option iii] was considered impractical because of the large number of different THR interventions in the NJR data base. Also for options iii] and iv] the number of recipients within some individual THR categories was too small to allow age and gender 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 iii] and iv] were considered less appropriate than options v] and vi].

For these two important reasons we therefore selected option v] for the base case. ***This represents a departure from the comparison specified in the protocol and scope.***
We selected option v] 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). (Figure 22 and Figure 39). We used propensity matching to match NJR patients with RS patients for objective A (see Chapter 7). Propensity matching by age and gender was undertaken using the *Edwin Leuven* procedure.[345]

Comparison of revision rates amongst these matched individuals was used in the economic analysis. We undertook subgroup analyses in which the RS versus THR comparison was examined separately for each gender 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 rate. There were three reasons for undertaking subgroup analyses: i) the difference between the genders in mechanical load bearing through the hip joint; [354] ii) the large difference in observed revision rates between men and women (see Section 9.2.2); and iii) expert clinical opinion which indicated that age represents a reasonable proxy for activity levels.

In the selection of alternative interventions to address our objective B (comparison between different types of THR), we were guided by the frequency of use of different prostheses and by clinical advice (see Chapter 7). 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. While 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 older than 65 years (Figure 23), 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 (< 65 years old) 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 gender).

However, because of the wide range of ages which received THR, and the different proportions of males and females receiving the different types of THR, we conducted sensitivity analysis controlling for age and gender. In addition, since only ~ 10% of RS recipients were > 65 years of age 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 (greater or less than 65 years) and were examined separately by gender. 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 less than 65 years, and for men and women aged 70 and 80 years using the population aged more than 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 NICE's previous technology assessments of hip replacement interventions (TA 2 and TA 44), which recommended, should evidence allow, that different interventions should be compared in subgroups of patients according to age and gender.[355] However, these subgroup analyses represent an extension from our protocol and scope. Table 57 summarises the makeup of THR population by age and gender.

Table 57. Age and gender characteristics of patient groups receiving THR prostheses

Population	Number	% Female	Mean age (SD)	Median	Inter quartile (IQ) range
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 CAT A to E recipients	239,089	63.5	71.6 (9.6)	72.5	65.8-78.3
All THR CAT A to E female recipients	151,902	100	72.1 (9.6)	73	66.4-78.9
All THR CAT A to E male recipients	87,187	0	70.5 (9.6)	71.5	64.9-77.1
All CAT A recipients	125,285	66.9	74.6 (7.9)	74.9	69.7-80
All CAT B recipients	37,874	60.2	71.5 (8.7)	72	65.9-77.5
All CAT C recipients	34,754	55.4	61.6 (9.9)	62.3	55.9-67.9
All CAT D recipients	28,471	64.2	73.0 (8.3)	73.4	67.8-78.7
All CAT E recipients	12,705	60.1	66.2 (9.6)	66.3	60.7-72.5
All CAT A male recipients	41,472	0	73.9 (7.7)	74.2	69.2-79.0
All CAT B male recipients	15,055	0	70.9 (8.6)	71.6	65.6-76.7
All CAT C male recipients	15,502	0	61.6 (9.8)	62.5	56-67.9
All CAT D male recipients	10,181	0	72.5 (8.1)	72.9	67.6-77.9
All CAT E male recipients	4,977	0	65.5 (9.4)	65.6	60.3-71.6
All CAT A female recipients	83,813	100	74.9 (8.0)	75.3	70.0-80.5
All CAT B female recipients	22,819	100	71.8 (8.8)	72.3	66.2-78
All CAT C female recipients	19,252	100	61.6 (9.9)	62.2	55.8-67.9
All CAT D female recipients	18,290	100	73.3 (8.5)	73.7	67.9-79.2
All CAT E female recipients	7,728	100	66.7 (9.7)	66.8	60.9-73.1

9.2.3 Overall revision rates, competing risks (CR) and rationale for analysis

Revision rates amongst NJR patients have been the subject of several recent publications.[15,16,308,343] Some investigators have used KM analysis while others have employed CR analysis in which the

event of interest is revision, and death is taken as a competing risk. In KM 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 19 for results). RS revision rate estimates were very similar for both KM and CR analyses and were similar to those reported by Smith et al. (2012).[15] For THR the KM analysis generated somewhat higher rates of revision than CR analysis.

Both KM and CR estimated revision rates were higher for females than males for RS and female revision rates were less than those for males for THR. For this reason some sensitivity analyses in the economic analyses that follow have been stratified according to gender. So as to remain consistent with all previous economic analyses of hip replacement technologies we have used the revision estimates from KM analysis, together with parametric modelling to predict the rate of revision beyond the observed data.

In practice several parametric models fitted the Kaplan-Meier estimates of revision well. However, on extrapolation the models generated quite different revision rates mainly determined by different modelled hazard during the extrapolation period, some models predicting increasing hazard (e.g. bath tub), others decreasing hazard (e.g. lognormal); an example is shown in Figure 25. Increasing hazard of revision 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 (see Figure 26).

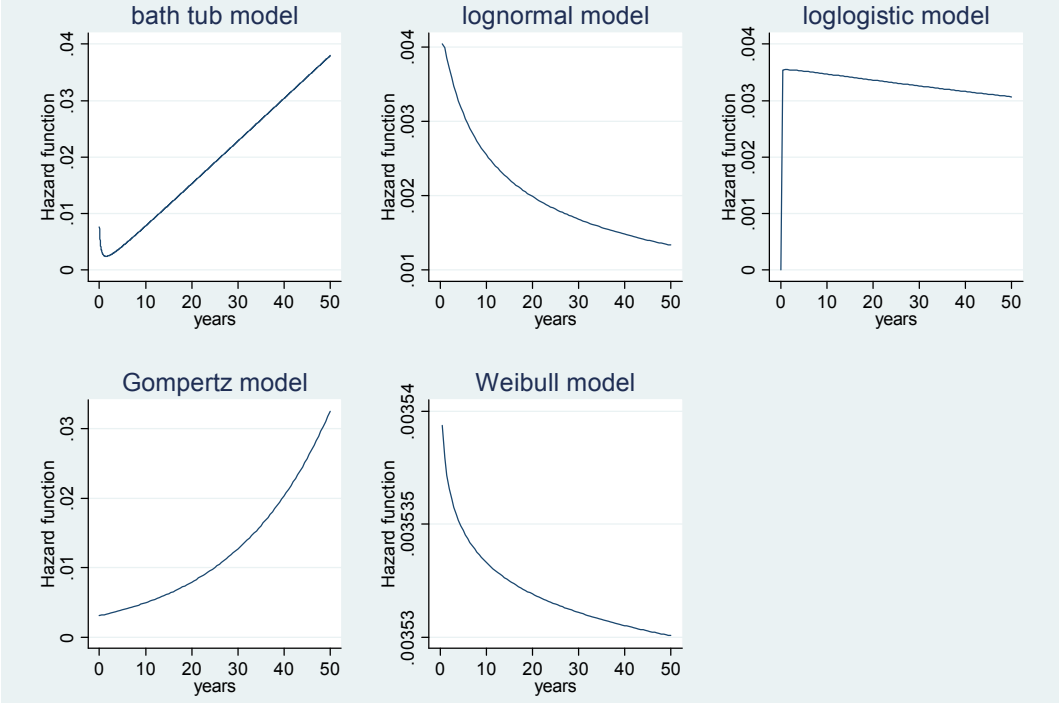


Figure 25. Differing modelled hazard in extrapolation beyond observation (HyPoM [Category D] THR females <65 years old)

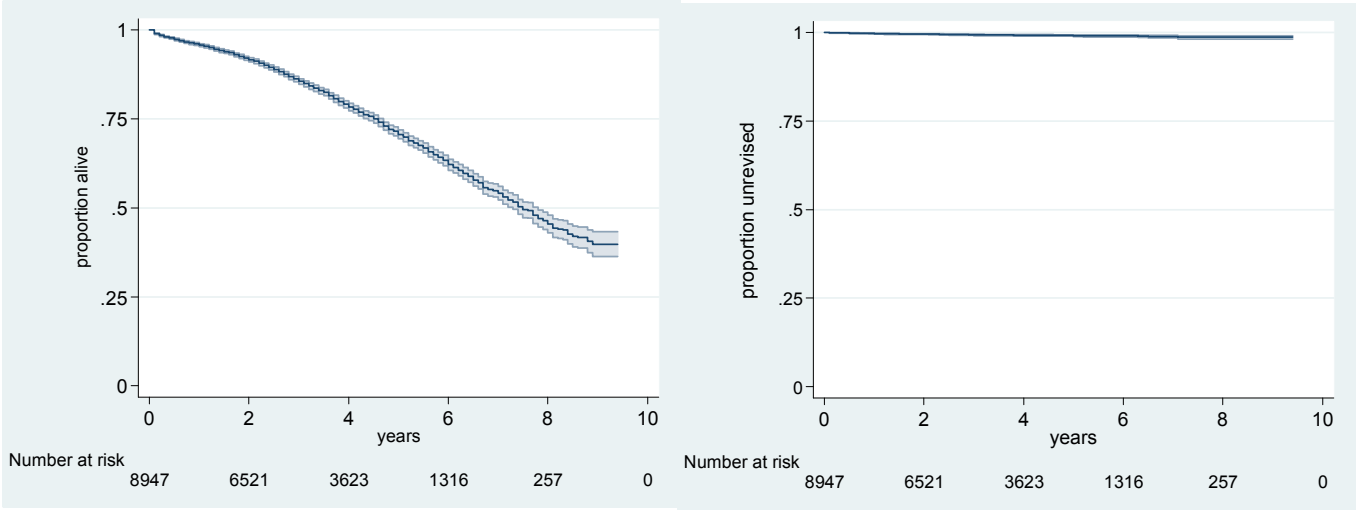


Figure 26. Kaplan Meier analysis for death (left) and revision (right) for THR CePoM (Category A) female patients > 85 years old

In view of these considerations, for the base case analysis we selected the best fit to the observed data across all the interventions which we compared. Since in practice the best fit was usually provided by the bath tub model (increasing hazard on extrapolation) sensitivity analyses were conducted with the best alternative fit which allowed for decreasing extrapolated hazard. In subgroup analyses according to age and gender 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. However, this guidance 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.[356,357] The NJR comprises observational rather than RCT data so that parametric fits for different interventions and or patient groups may not be well described by a single parametric form. Published cost-effectiveness analyses of hip replacement, have predominantly adopted a bath tub hazard model for revision rates.[38,44,272,358]

Information criteria (AIC, Bayesian Information Criterion (BIC)) scores for modelled fits, and plots of modelled log cumulative hazard versus log KM estimated hazard were used to judge goodness of fit and are provided in the main text or in Appendix 20.

9.3 Results

Parametric modelling results are reported in full in Appendix 18.

9.3.1 Proportional hazards tests

The condition of proportional hazards between observed revision rates for compared groups was examined using log KM estimated cumulative hazard versus log time. The results for RS versus THR and for the five categories of THR prosthesis are shown in Figure 27 and Figure 28 and Figure 29 respectively.

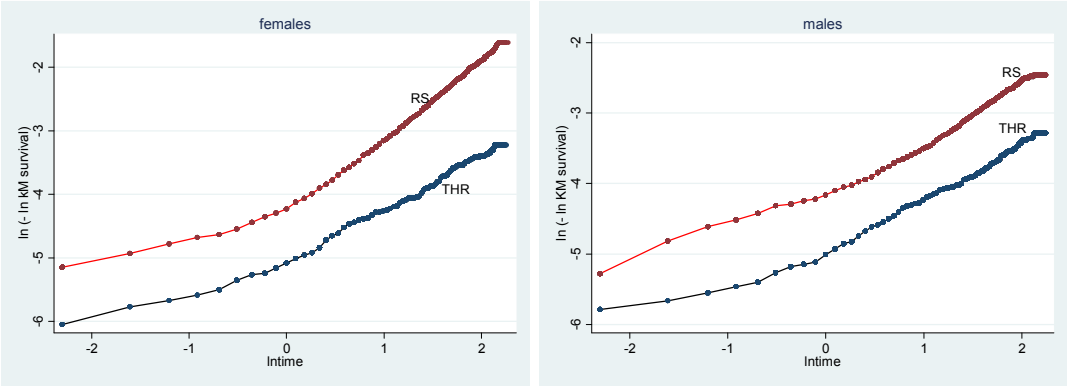


Figure 27. Log KM estimated cumulative hazard versus log time.
RS =resurfacing; THR = total hip replacement

Cumulative hazard plots for women for the comparison RS versus THR are not parallel; for men a proportional hazards assumption appears to hold moderately well. For the comparison of different THR prostheses, again the cumulative hazards were not noticeably parallel (Figure 28) this held also for THR categories when the population was stratified by gender and age Figure 29. Since 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.

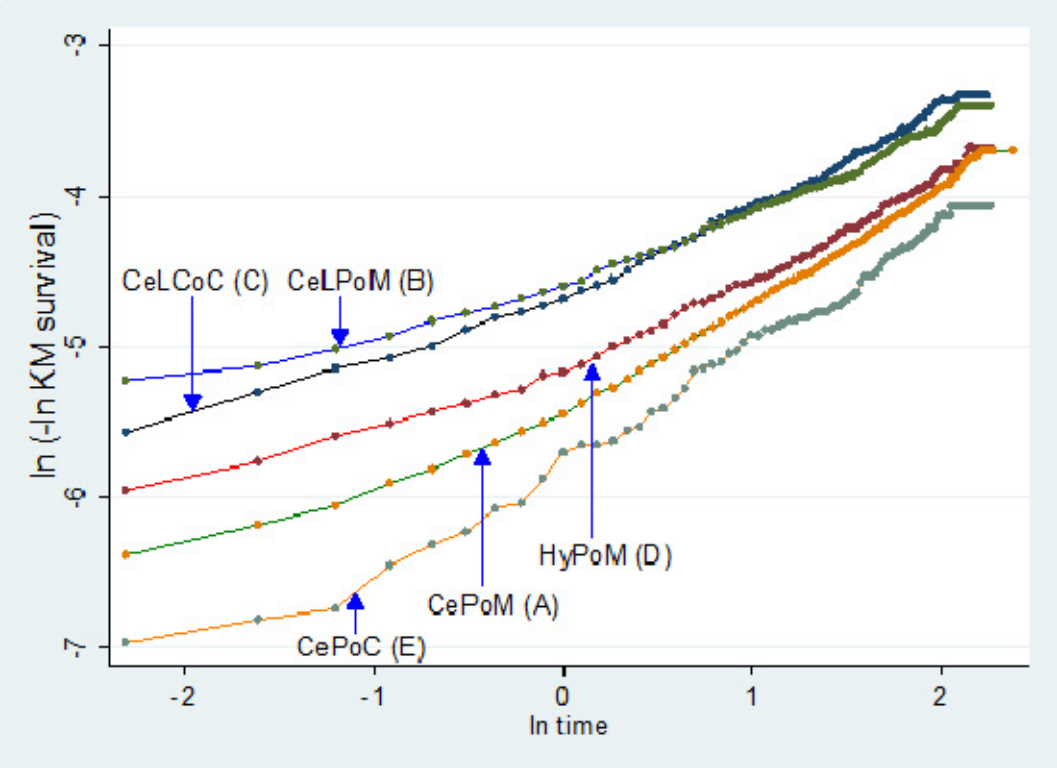


Figure 28. Log K-M estimated cumulative hazard versus log time for different THR categories

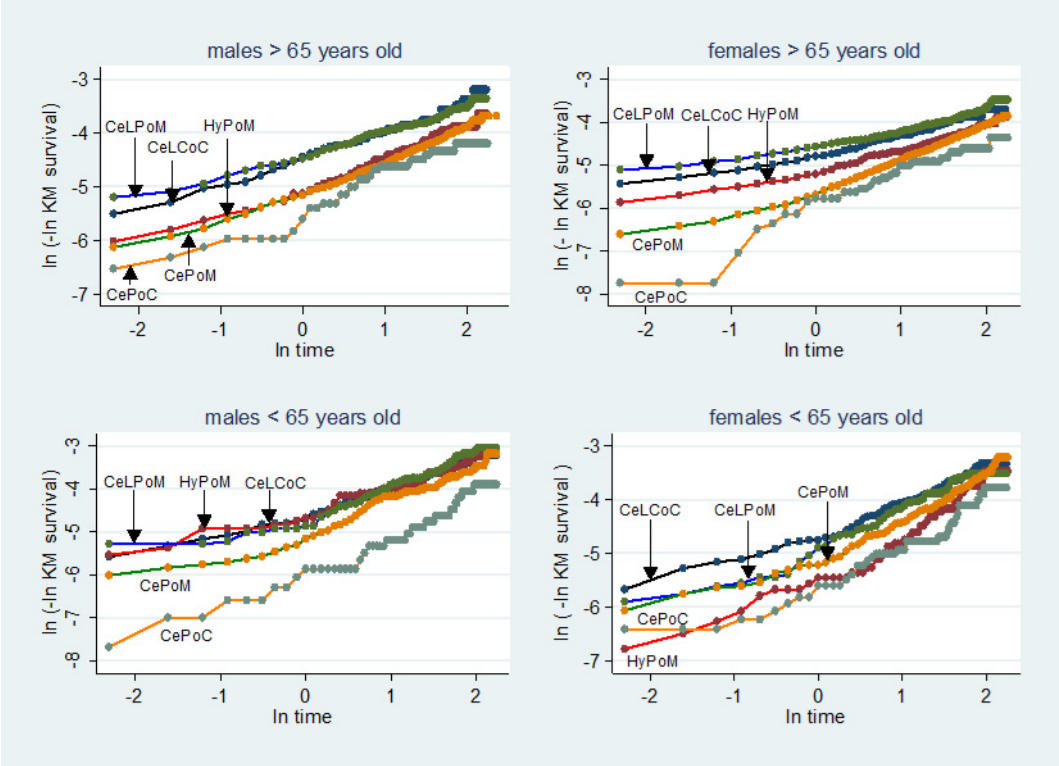


Figure 29. Log K-M estimated cumulative hazard versus log time for different THR categories

Populations stratified by gender and age.

9.3.2 Comparison of RS vs. THR

For both genders many more patients received THR than RS. The observed revision rate for all RS recipients (n=31,222) over nine years of follow up was about three times that for all THR recipients (n=386,556) (Figure 30 and Figure 31). When the comparison was made by gender the observed revision rate for female RS recipients was more than three times that of female THR patients and for male RS recipients about twice that for male THR recipients.

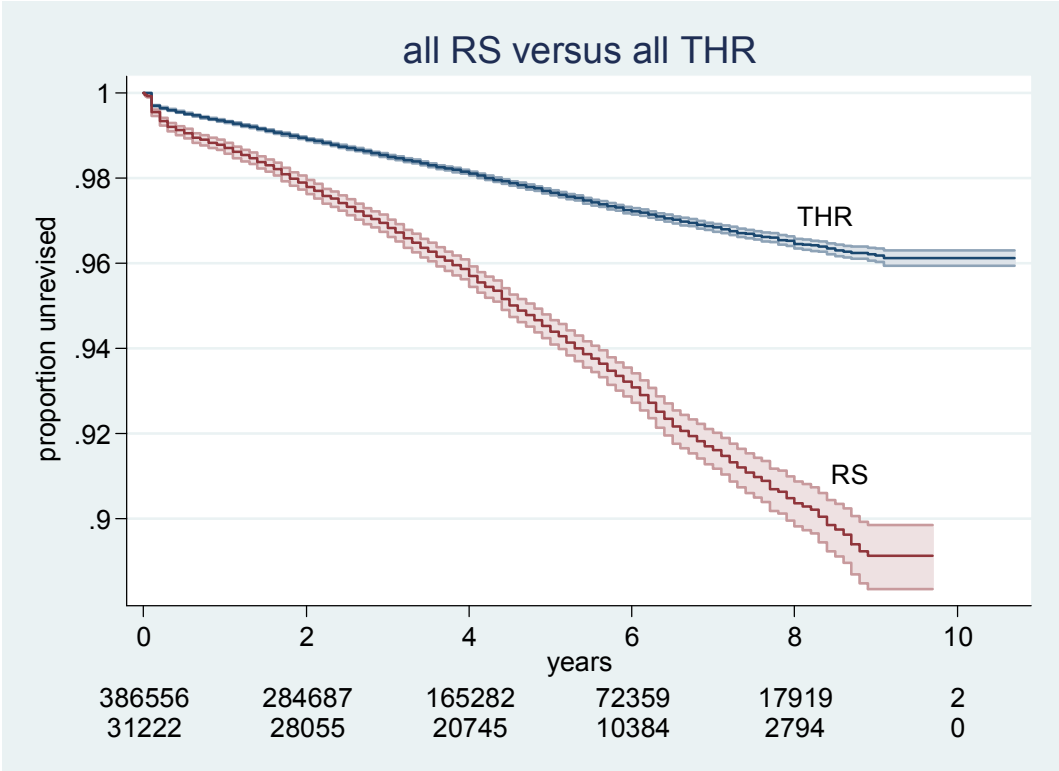


Figure 30. Time to revision; all RS patients and all THR patients
Numbers under x axis are numbers at risk (THR upper, RS lower)

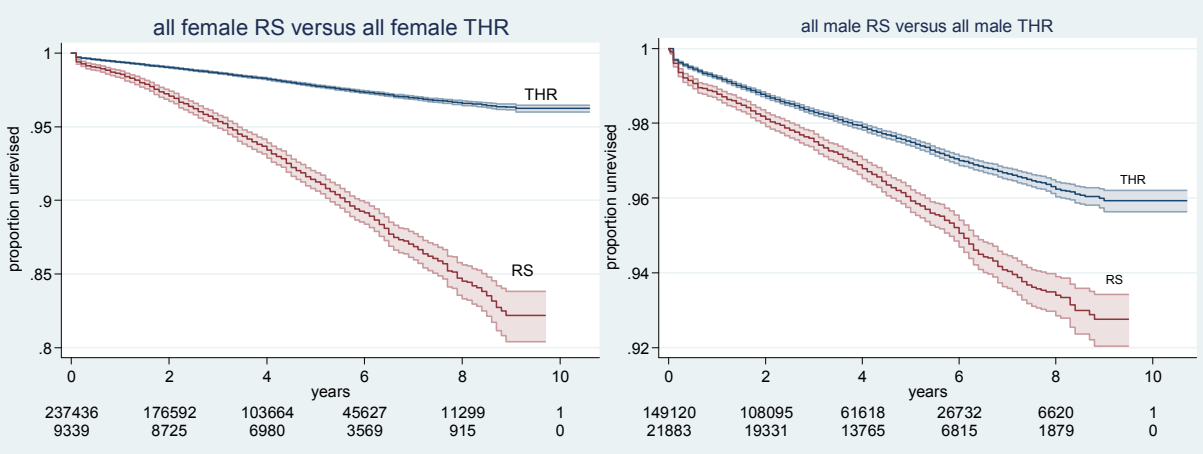


Figure 31. Time to revision; all RS and all THR patients according to gender

Numbers under x axis are numbers at risk: THR upper, RS lower

When the RS versus THR comparison was restricted to THR recipients of the five prosthesis categories A to E (n=239,089) the differences were larger (Figure 32) and again held across gender. When revision rates for recipients of the individual categories of THR were compared with all RS recipients the observed revision rates for both genders were considerably higher for RS than for any single THR category.

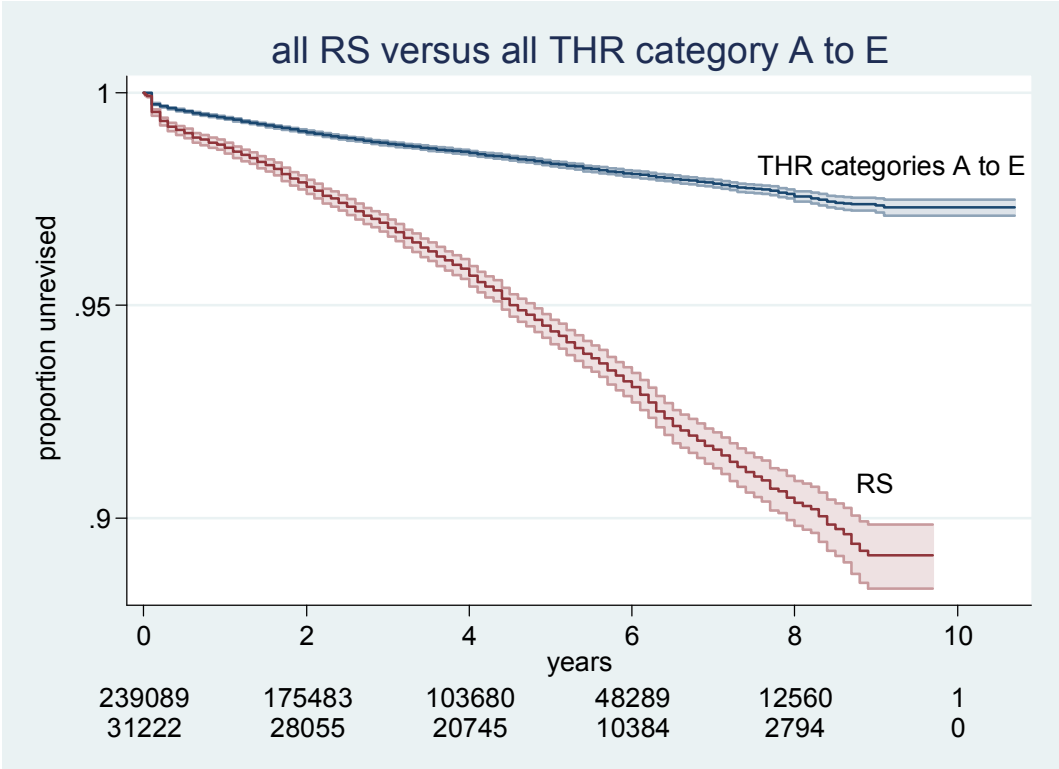


Figure 32. Time to revision all RS and all THR patients (Categories A to E)
Numbers under x axis are numbers at risk: THR upper, RS lower

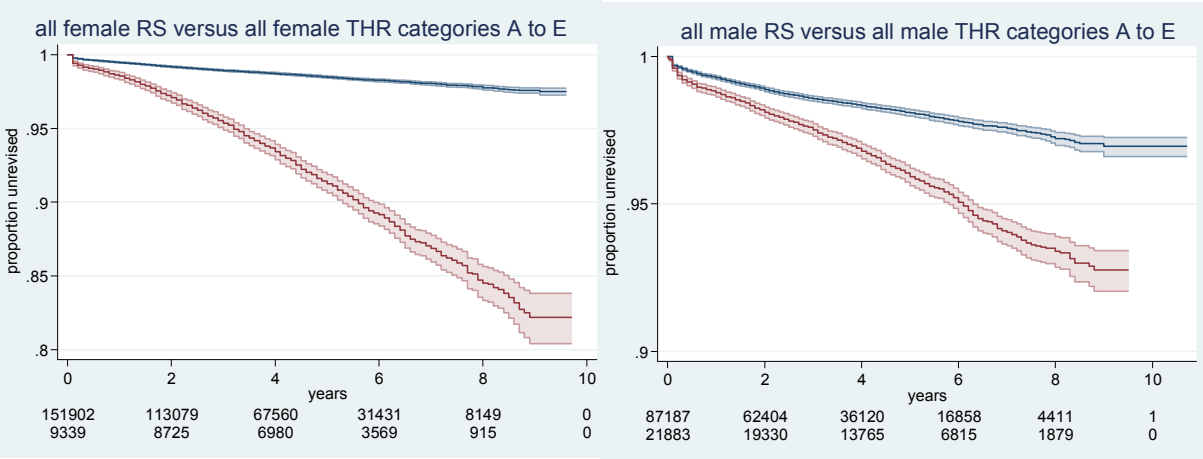


Figure 33. Time to revision all RS and all THR patients (Categories A to E) by gender
Numbers under x axis are numbers at risk: THR upper, RS lower

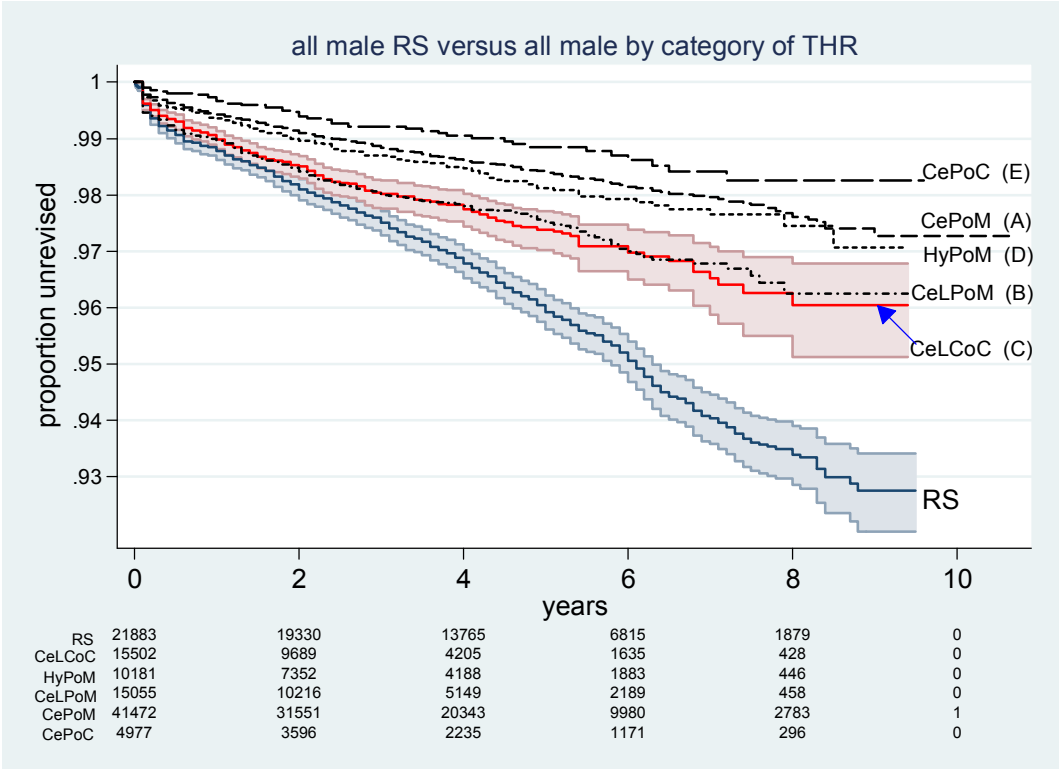


Figure 34. Time to revision all male RS and all male THR patients (Category A to E)
Numbers under x axis are numbers at risk

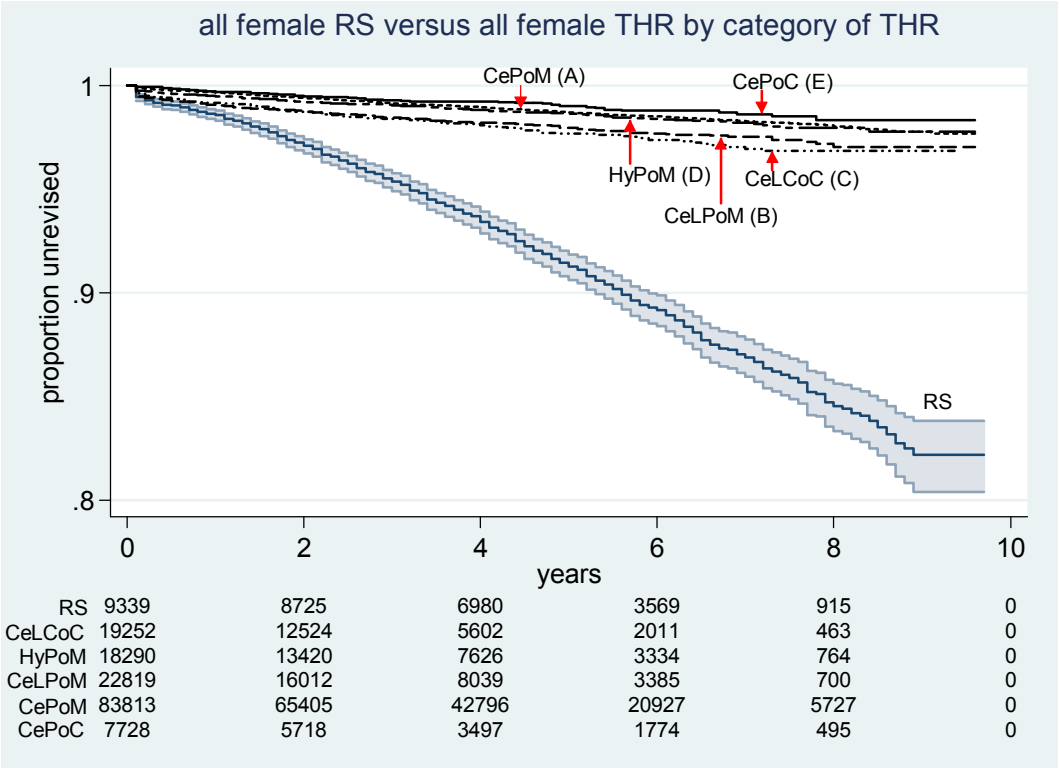


Figure 35. Time to revision all female RS and all female THR patients (Category A to E)

Numbers under x axis are numbers at risk

Table 58. Age and gender mix of RS and THR populations

Population	Number	%Female	Mean age (SD)	Median	IQ range
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 Categories 1 to 5 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	9,321	100	53.5 (8.4)	54.0	49-59
THR propensity matched population female	9,321	100	53.5 (8.4)	54.0	49-59

It is clear that revision rates after RS are much higher for both genders than those after THR of any category. However, age and gender difference between the RS and THR populations (see Table 58. Age and gender mix of RS and THR populations make these comparisons inequitable. More males than females received RS while more females than males received THR, and nearly all RS recipients were aged less than 65 years (mean age ~56 years) whereas most THR recipients were more than 65 (mean ~72years). For an equitable comparison of the interventions is necessary to match populations by gender and age.

Of the male and female patients who received the RS intervention for osteoarthritis 17,322 and 9,321 respectively were successfully propensity matched by age with THR patients from THR categories A to E (n=239,089), providing 26,643 matched pairs for comparison (see Chapter 7 on matching). Age distribution was identical in the RS and THR matched populations (Table 58) but was slightly skewed from normal (Figure 36). KM analysis (Figure 36) revealed that revision rates were much higher for RS than for the matched THR population.

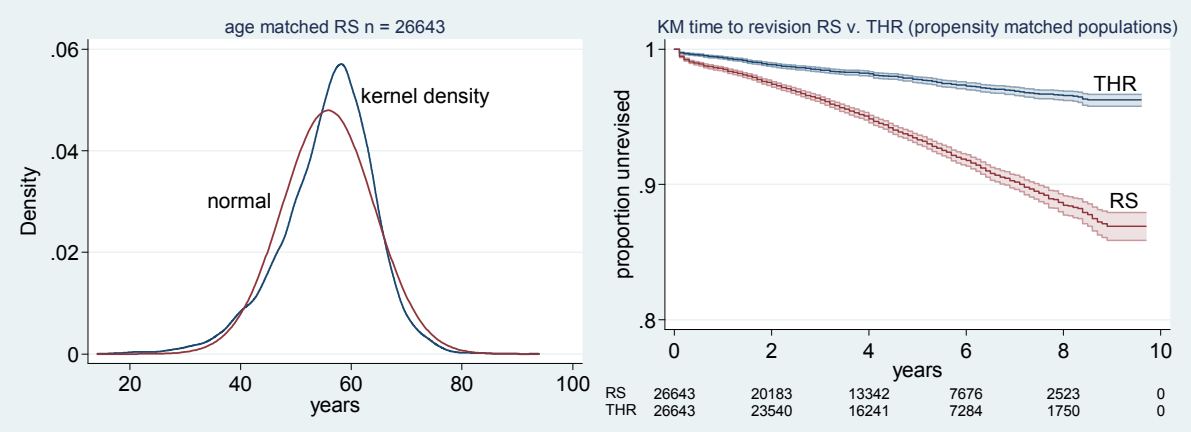


Figure 36. Age distribution and Time to revision for RS and THR matched populations

Numbers under x axis are numbers at risk

Revision was more frequent amongst the matched THR population than the whole THR population (Figure 37) demonstrates the importance of the matching process prior to comparison of RS versus THR.

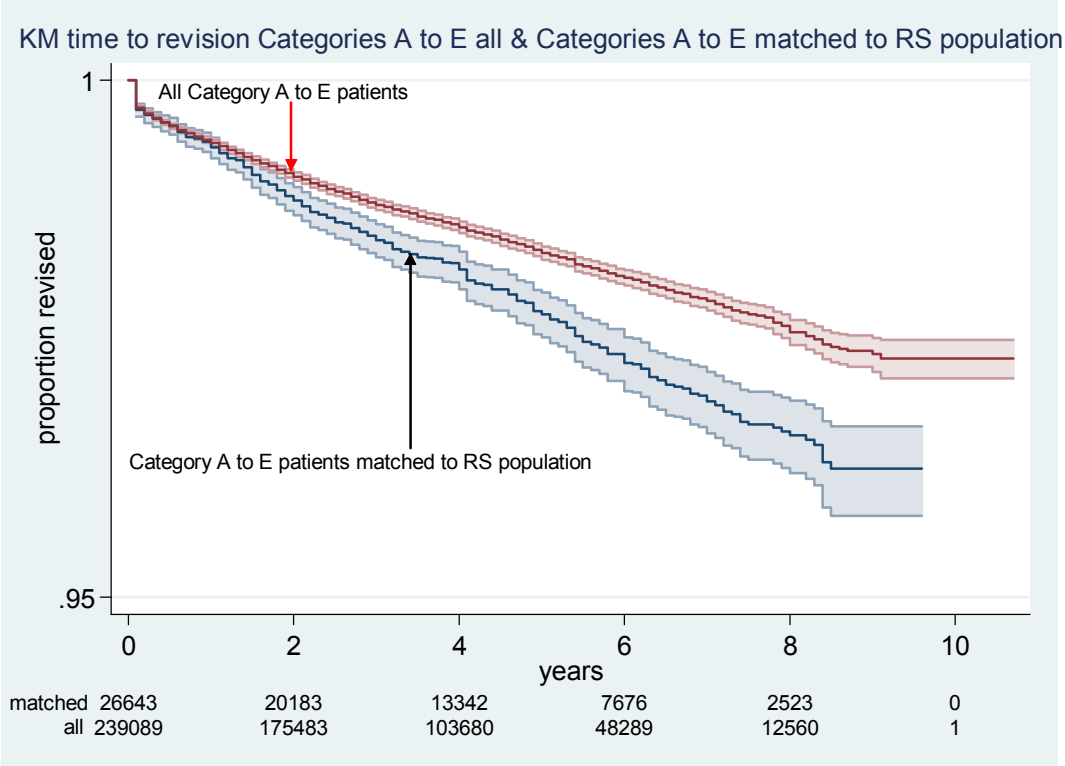


Figure 37. Revision rates for matched and whole THR populations

Numbers under x axis are numbers at risk

Information criteria (Table 59) indicated that bath tub models provided the best fit for both RS and THR shown in Figure 36. Therefore for comparing RS with THR in the base case economic analysis, transition probabilities were calculated using bath tub model. Bathtub fits and extrapolations are shown in Figure 38 and reflect clinical practice as represented by patients in the NJR database. Bath tub fits were supported visually (Appendix 18) and by plots of modelled versus KM-estimated cumulative hazard (Figure 39).

Table 59 Information criteria scores for models of revision rates (RS and matched THR)

Intervention	Model	Observations	Model likelihood	Parameters	AIC	BIC
THR	exponential	26643	-3239.377	1	6480.753	6488.944
THR	Weibull	26643	-3219.967	2	6443.935	6460.315
THR	Gompertz	26643	-3230.912	2	6465.825	6482.205
THR	lognormal	26643	-3221.913	2	6447.827	6464.207
THR	loglogistic	26643	-3220.111	2	6444.222	6460.603
THR	bath tub	26643	-3215.51	3	6437.021	6461.592
RS	exponential	26643	-8102.451	1	16206.9	16215.09
RS	Weibull	26643	-8101.688	2	16207.38	16223.76
RS	Gompertz	26643	-8094.569	2	16193.14	16209.52
RS	lognormal	26643	-8162.981	2	16329.96	16346.34
RS	loglogistic	26643	-8107.527	2	16219.05	16235.43
RS	bath tub	26643	-8037.685	3	16081.37	16105.94

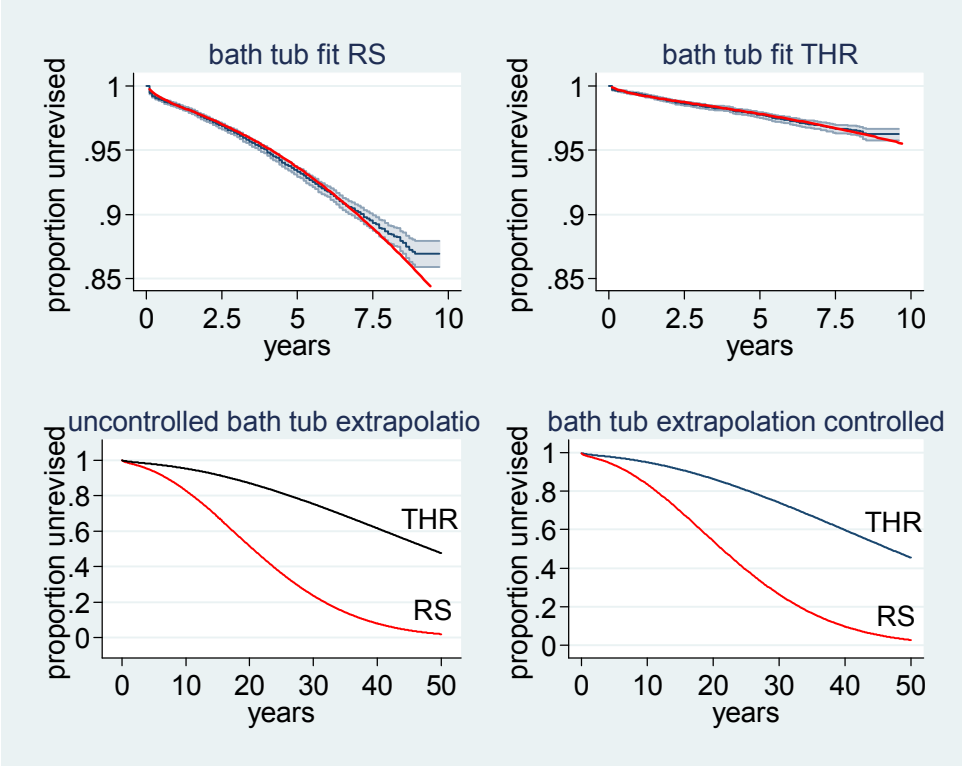


Figure 38. Bath tub fits and extrapolations for matched THR and RS populations

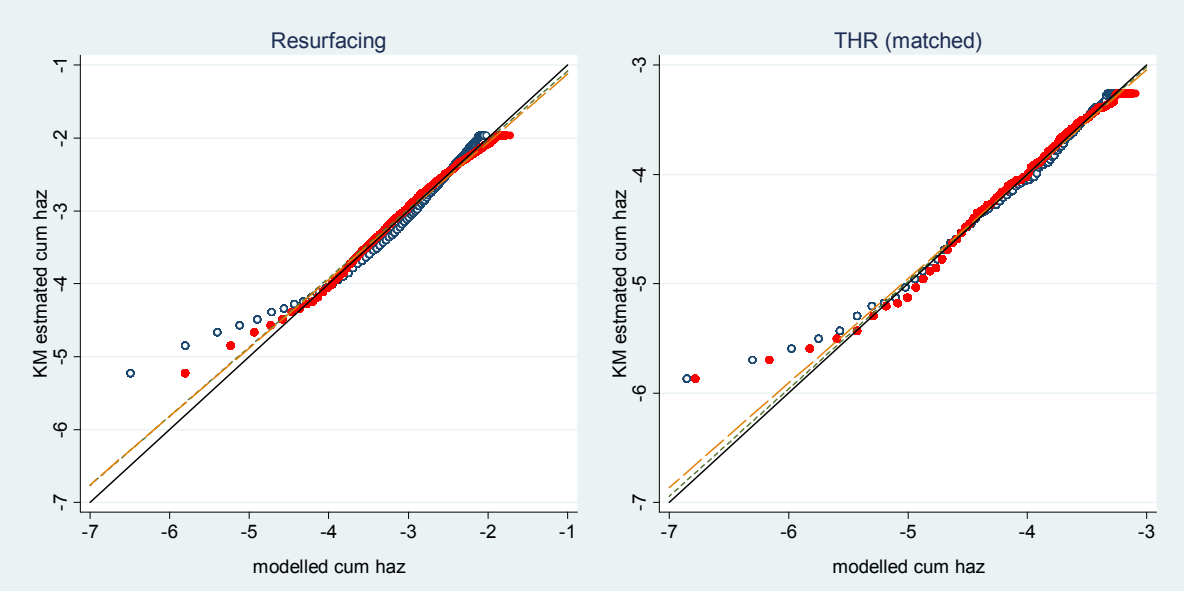


Figure 39. K-M versus modelled cumulative hazard
Open symbols loglogistic fit closed symbols bath tub fit; Lines: solid= KM; short dash = linear regression for bathtub; long dash= linear regression for loglogistic

Table 60. Bath tub modelled percentage of patients requiring revision

Intervention	10 years	20 years	30 years
RS	17.2	48.3	76.3
THR	4.6	12.9	24.6

Bath tub modelled percentage revision at 10, 20 and 30 years is summarised in Table 60.

Since the age distributions of the matched populations were somewhat removed from normal (Figure 36) we undertook sensitivity analysis in which bath tub models were controlled for age and gender and extrapolated revision calculated for an "average" population of 35% female aged 55.8 years (Figure 38). Because it was evident that revision rates were much higher for women receiving RS than for men, and because revision rates likely vary according to the age of patients, subgroup analyses focussed on comparing populations stratified by gender and controlled for age. The results of the analysis of revision rates for these subgroups are provided in following sections and in Appendix 18.

9.3.3 Comparison of THR categories

There were recipients of whom were encompassed within THR categories at nine years the KM estimated proportion remaining unrevised was 0.974 for the 239,089 patients from the five selected THR categories (A to E). and 0.962 for all 386,556 NJR THR recipients (Figure 40). The KM plot for the five selected THR interventions indicated a relatively high initial hazard for revision which gradually decreased over about four years and subsequently gradually increased between five and nine years.

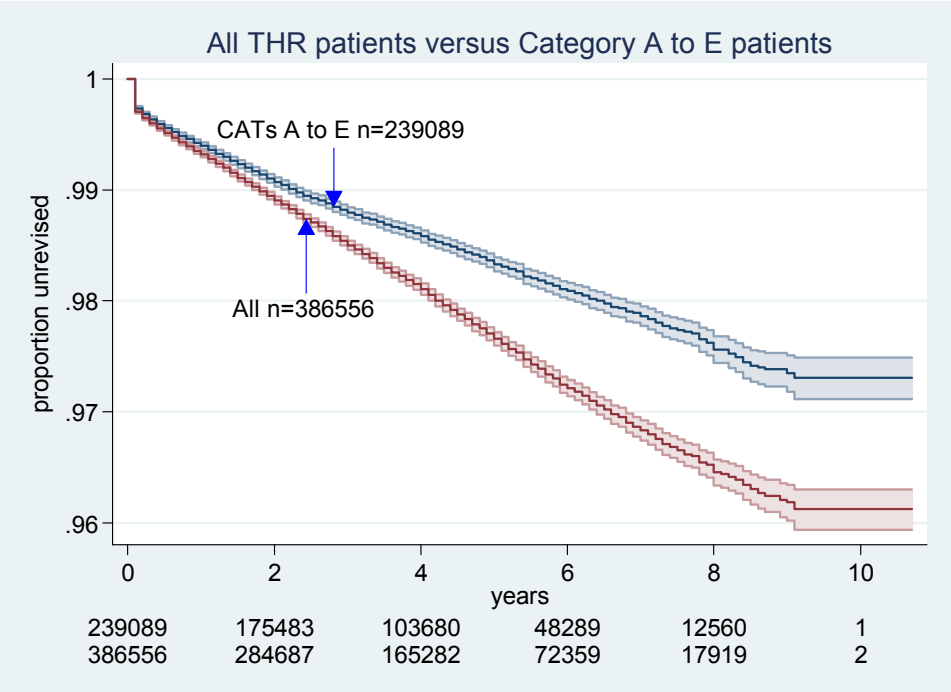


Figure 40. Revision estimated for all THR patients and those receiving category A to E THRs

Numbers under x axis are numbers at risk

KM analyses indicated different revision rates across the five categories of THR **Error! Reference source not found.**). Revision rates for patients who received CeLCoC (C) and CeLPoM (B) THRs were clearly higher than those who received CePoC (E) and CePoM (A).

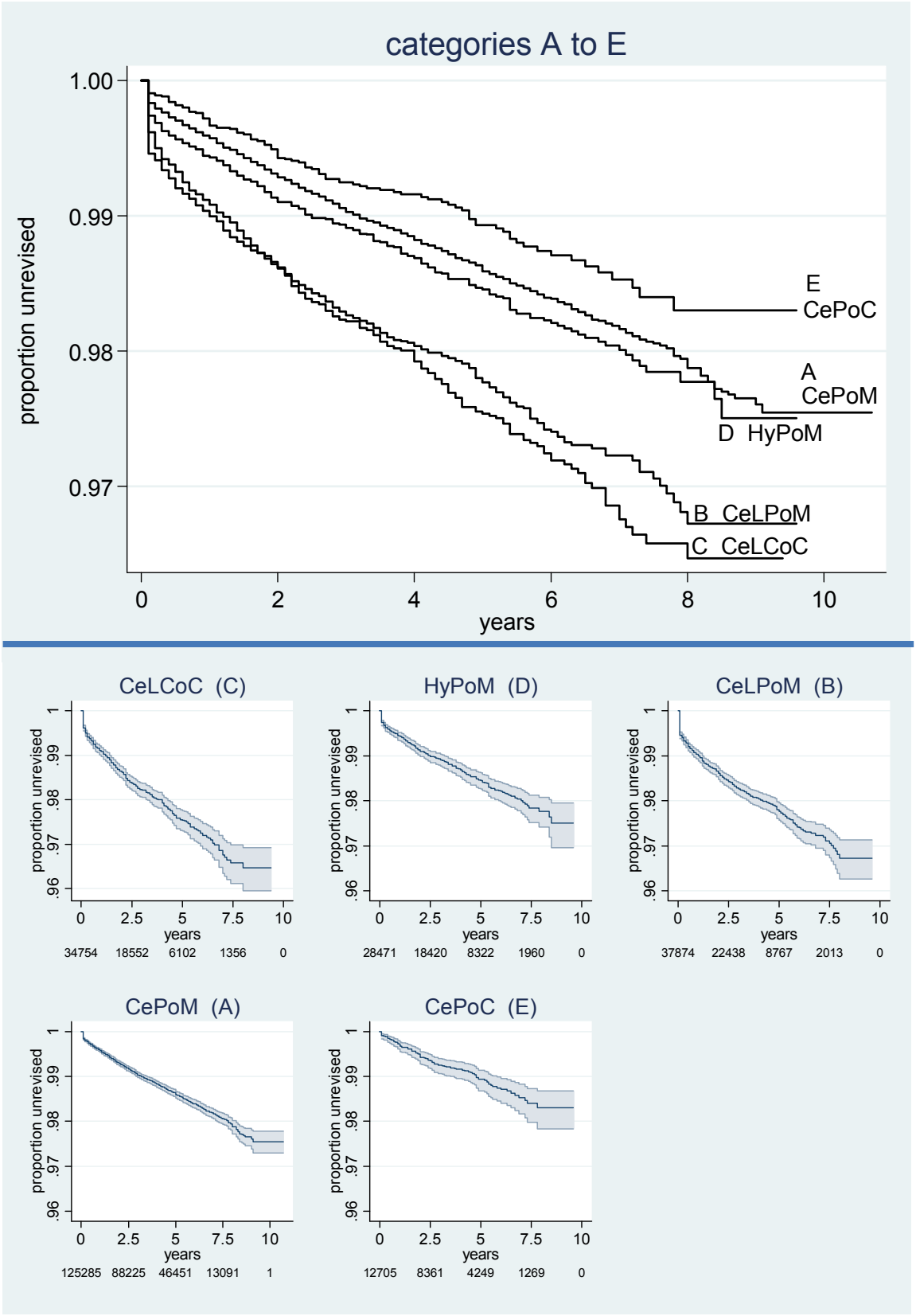


Figure 41. Observed time to revision; upper panel THR categories compared; lower panel THR categories with 95% CI; cement-less upper row, cemented lower row

Numbers under x axis refer to numbers at risk

Table 61. Information criteria scores for models of revision rates (THR categories)

THR	Model	observations	Model likelihood	Parameters	AIC	BIC
CeLCoC (C)	exponential	34754	-3955.734	1	7913.467	7921.923
CeLCoC (C)	Weibull	34754	-3882.115	2	7768.229	7785.141
CeLCoC (C)	Gompertz	34754	-3906.282	2	7816.563	7833.475
CeLCoC (C)	lognormal	34754	-3872.162	2	7748.323	7765.235
CeLCoC (C)	loglogistic	34754	-3881.911	2	7767.822	7784.734
CeLCoC (C)	bath tub	34754	-3858.878	3	7723.755	7749.123
HyPoM (D)	exponential	28471	-2428.234	1	4858.468	4866.724
HyPoM (D)	Weibull	28471	-2387.427	2	4778.854	4795.368
HyPoM (D)	Gompertz	28471	-2405.936	2	4815.872	4832.385
HyPoM (D)	lognormal	28471	-2383.97	2	4771.94	4788.454
HyPoM (D)	loglogistic	28471	-2387.411	2	4778.822	4795.335
HyPoM (D)	bath tub	28471	-2373.646	3	4753.291	4778.061
CeLPoM (B)	exponential	37874	-4535.478	1	9072.955	9081.497
CeLPoM (B)	Weibull	37874	-4391.882	2	8787.763	8804.847
CeLPoM (B)	Gompertz	37874	-4442.601	2	8889.202	8906.286
CeLPoM (B)	lognormal	37874	-4377.507	2	8759.014	8776.098
CeLPoM (B)	loglogistic	37874	-4391.567	2	8787.133	8804.217
CeLPoM (B)	bath tub	37874	-4345.8	3	8697.601	8723.227
CePoM (A)	exponential	125285	-10000.51	1	20003.01	20012.75
CePoM (A)	Weibull	125285	-9929.73	2	19863.46	19882.94
CePoM (A)	Gompertz	125285	-9965.745	2	19935.49	19954.97
CePoM (A)	lognormal	125285	-9927.767	2	19859.53	19879.01
CePoM (A)	loglogistic	125285	-9929.867	2	19863.73	19883.21
CePoM (A)	bath tub	125285	-9909.508	3	19825.02	19854.23
CePoC (E)	exponential	12705	-759.4492	1	1520.898	1528.348
CePoC (E)	Weibull	12705	-757.1662	2	1518.332	1533.232
CePoC (E)	Gompertz	12705	-757.8727	2	1519.745	1534.645
CePoC (E)	lognormal	12705	-756.8497	2	1517.699	1532.599
CePoC (E)	loglogistic	12705	-757.163	2	1518.326	1533.226
CePoC (E)	bath tub	12705	-756.6023	3	1519.205	1541.554

According to information criteria (Table 61), other than for CePoC (E), the bath tub model provided the best parametric fit, followed by the lognormal. For CePoC (E) lognormal was marginally superior to bath tub. These inferences were supported by visual inspection (Appendix 18) and by comparing modelled with KM estimated cumulative hazards for each category (Figure 42).

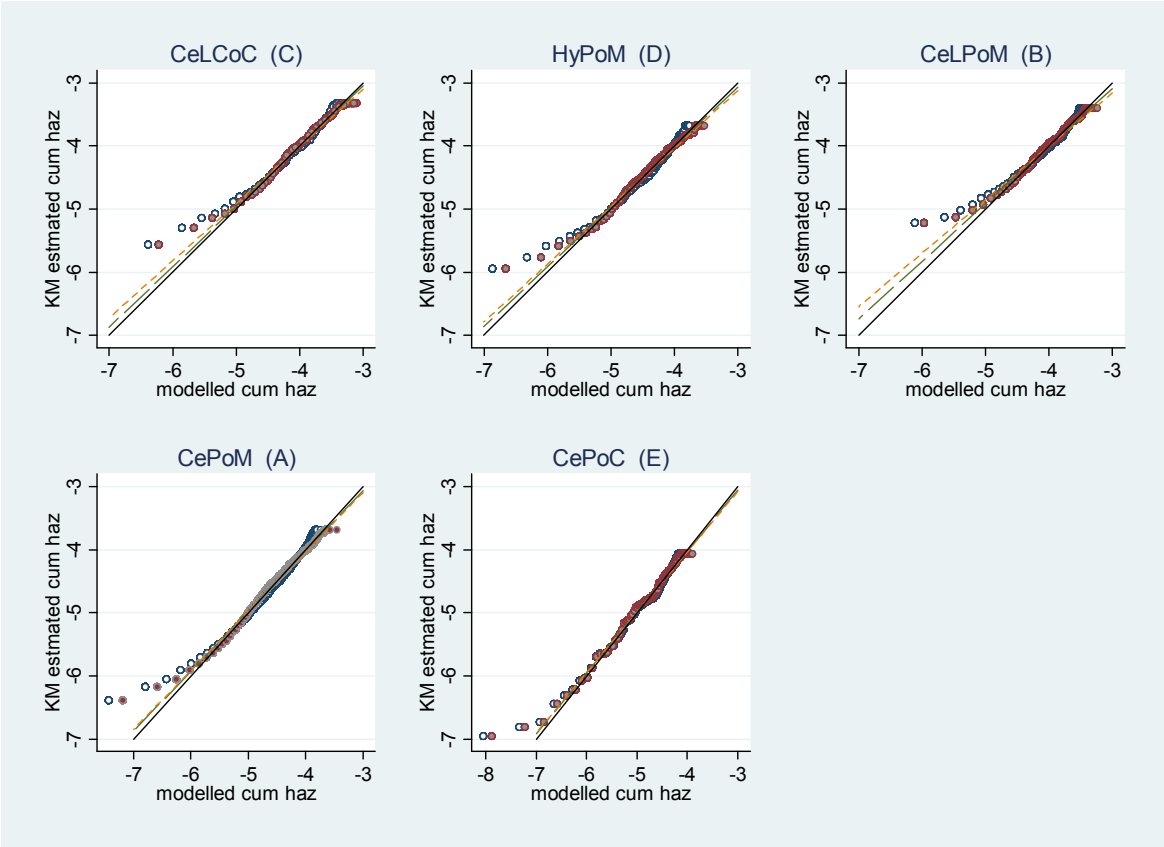


Figure 42. K-M versus modelled cumulative hazard

Cement-less upper row, cemented lower row; open symbols lognormal fit closed symbols bath tub fit; Lines: solid= KM; short dash = linear regression for bathtub; long dash= linear regression for lognormal.

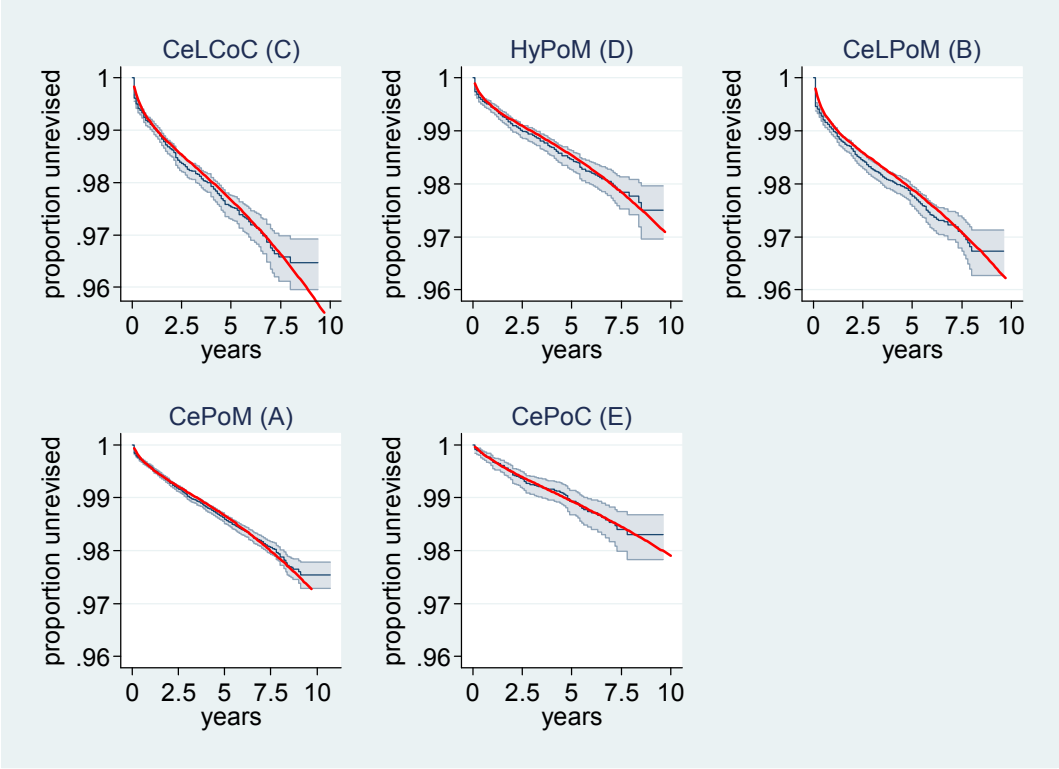


Figure 43. Bath tub parametric fits to observed time to revision for THR categories A to E

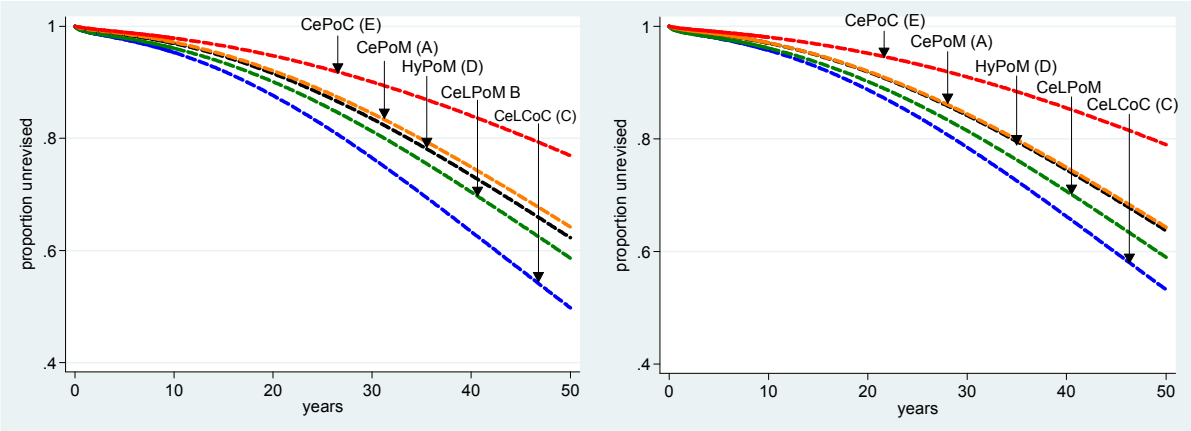


Figure 44. Extrapolation of bath tub models of revision for THR categories A to E
Left uncontrolled. Right controlled for age and gender, modelled population aged 71.6 years, 63.5% female.

For the base case economic analysis, transition probabilities were calculated from the bath tub fits for each category A to E; the fit to the KM estimates and the extrapolation beyond the observed data are shown in Figure 43and Figure 44 (left) respectively. These analyses reflect the performance of the five types of prosthesis for NJR patients over nine to 10 years to 2012. The lowest and highest revision rates were experienced by CePoC (E) and CeLCoC (C) recipients respectively. The bath tub modelled percentage of patients requiring revision at 10, 20 and 30 years is summarised in Table 62.

Table 62. Bath tub modelled percentage of patients requiring revision

THR category	10 years	20 years	30 years
CePoM (A)	2.8	7.9	15.6
CeLPoM (B)	3.9	9.9	18.7
CeLCoC (C)	4.6	12.3	23.5
HyPoM (D)	3.0	8.4	16.5
CePoC (E)	2.1	5.2	9.9

Table 63. Age and gender of recipients of THR categories A to E

Population	Number	%Female	Mean age (SD)	Median	IQ range
All THR (CAT A-E) recipients	239,089	63.5	71.6 (9.6)	72.5	65.8-78.3
All CePoM (A) recipients	125,285	66.9	74.6 (7.9)	74.9	69.7-80
All CeLPoM (B) recipients	37,874	60.2	71.5 (8.7)	72	65.9-77.5
All CeLCoC (C) recipients	34,754	55.4	61.6 (9.9)	62.3	55.9-67.9
All HyPoM (D) recipients	28,471	64.2	73.0 (8.3)	73.4	67.8-78.7
All CePoC (E) recipients	12,705	60.1	66.2 (9.6)	66.3	60.7-72.5

Across the five THR categories there were 36.5% men and 63.5% women, but within categories the ratio varied from 1.24 for CeLCoC (C) to 2.02 for CePoM (Table 63). Revision was more frequent for men than women (Figure 45), although this was least pronounced for the CePoC prosthesis.

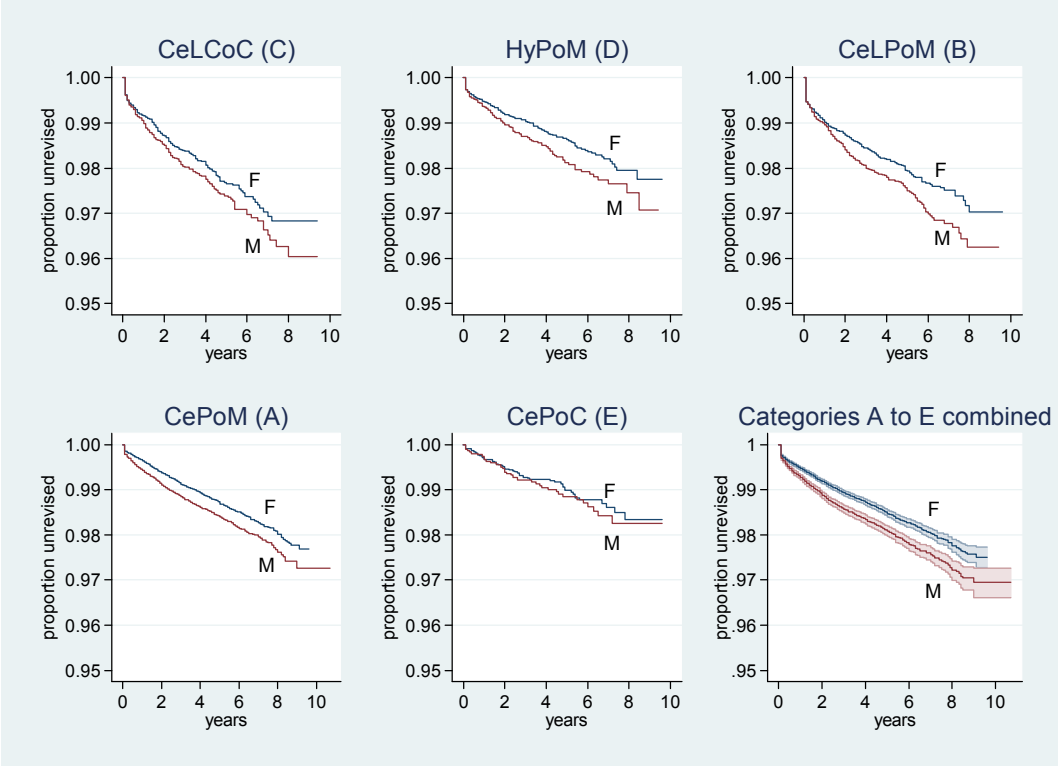


Figure 45. THR revision rates observed for males and females: all THR categories

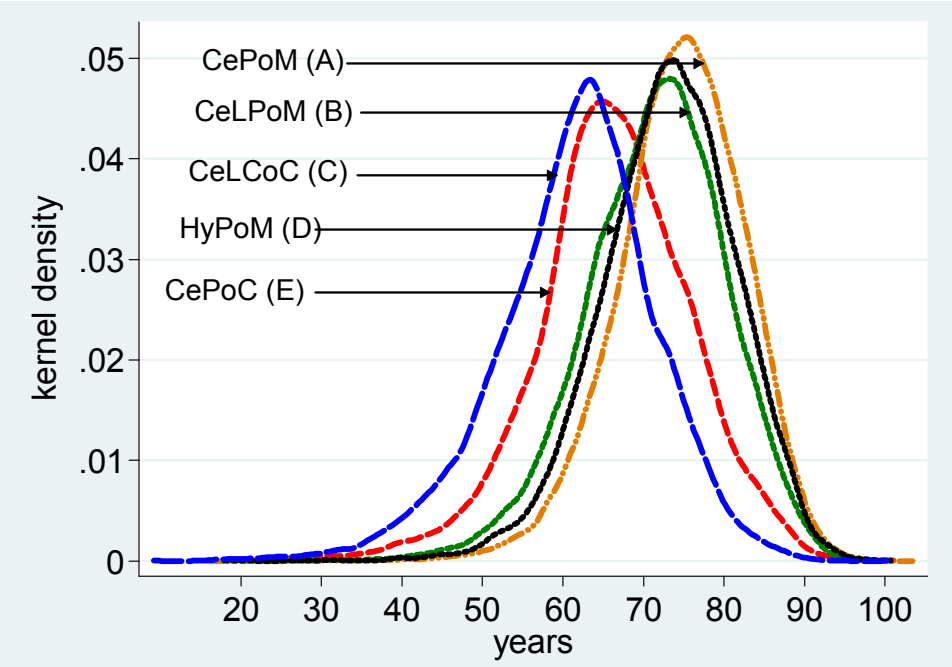


Figure 46. Kernel density plots of age at primary for category A to E THR prostheses

Similarly the age distribution of patients differed somewhat according to THR category (Figure 46), CeLCoC (C) was used more for younger patients and CePoM (A) for older (Table 63). Across the five THR categories the mean age was 71.56 years

In sensitivity analysis the bath tub model was controlled for age and gender, in order to adjust for spurious differences in revision rates due to differing proportions of men and women or of younger or older patients in THR categories. Figure 44 (right) shows this. The relative performance of the five categories modelled for populations aged 71.6 years, 63.5% female demonstrates that the superiority of the CePoC prosthesis was somewhat enhanced.

In further sensitivity analysis we used lognormal fits to the KM estimated revision; these are shown for each of the types of THR (categories A-E) in Figure 47. With a mean age across all categories of nearly 72 years, extrapolation predicting decreasing hazard for revision may be appropriate. The best fit model providing this condition was the lognormal. These fits are shown in Figure 48. The relative performance of the prostheses was similar to that with the bath tub model however, unsurprisingly; extrapolated revision rates were lower than with the bath tub model.

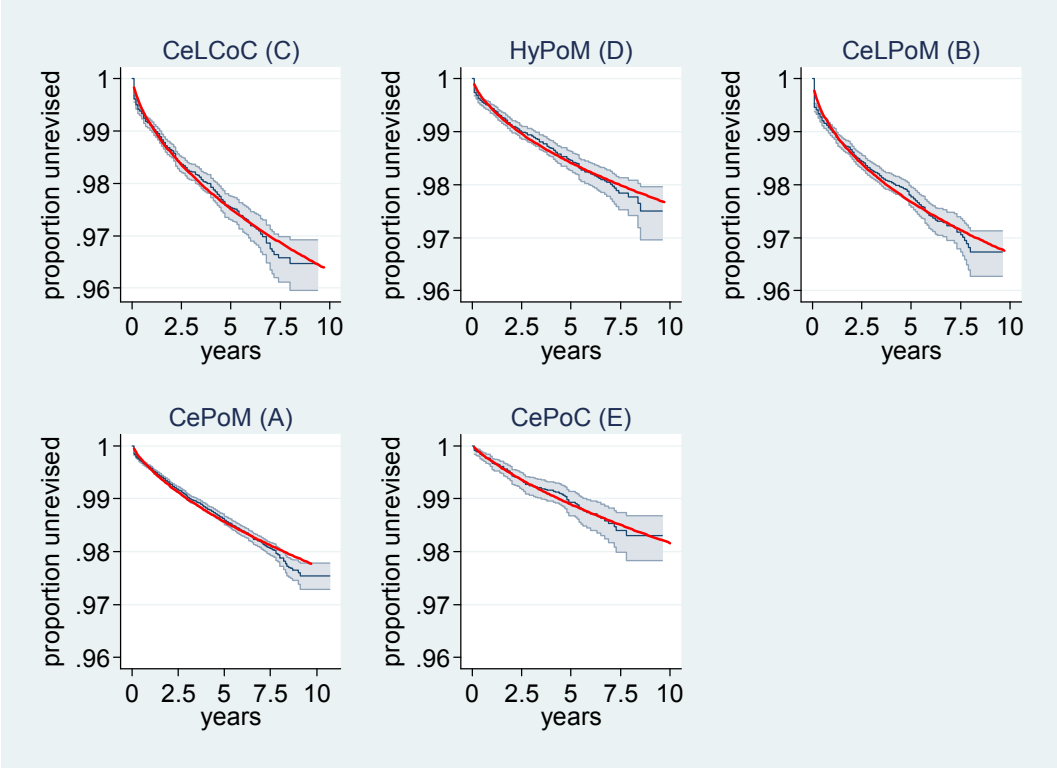


Figure 47. Lognormal parametric fits to observed revision for THR categories A to E

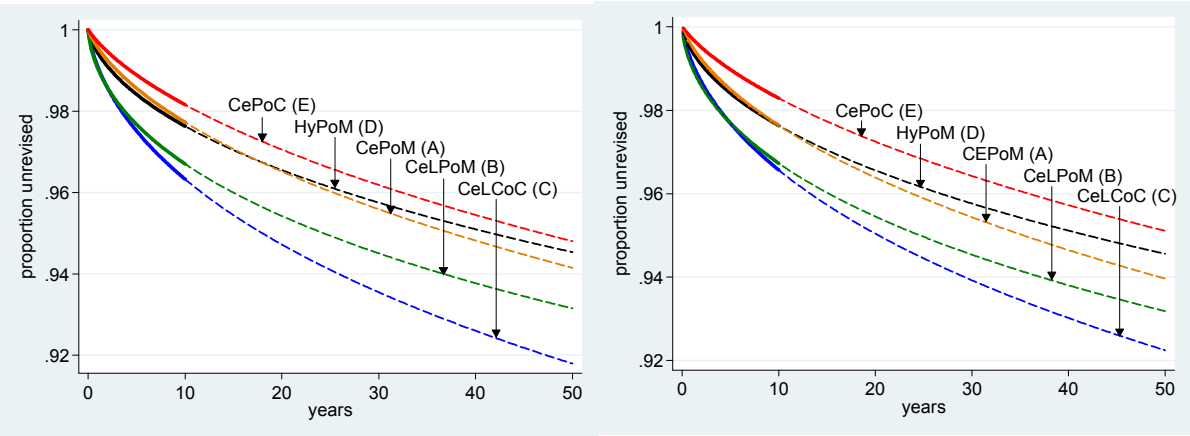


Figure 48. Lognormal modelled revision; uncontrolled (left), controlled for population of mean age 71.6 years and 36.5% male

Table 64. Lognormal modelled percentage of patients requiring revision

THR category	10 years	20 years	30 years
CePoM (A)	2.3	3.5	4.4
CeLPoM (B)	3.3	4.6	5.5
CeLCoC (C)	3.7	5.3	6.4
HyPoM (D)	2.4	3.4	4.2
CePoC (E)	1.8	2.9	3.8

A further sensitivity analysis was done in which the lognormal model was controlled for age and gender; with this model the superior performance of the CePoC (E) prosthesis was maintained (Figure 48, right).

9.3.4 Comparison of RS and THR; subgroup analyses according to gender (females)

Since the use of different categories of THR prostheses differed by age and gender and since recipients of hip replacement interventions aged > 65 years approximate a population unlikely to be considered candidates for RS (Figure 24) we undertook subgroup analyses in which the THR population for each category was stratified by gender and by age (> and < 65 years old), and parametric models were controlled for age. Results from these analyses are presented below.

As expected, the matched groups (n 9,321) had identical age distribution: mean age 53.5 years (SD 8.4; range 15 to 93) (Figure 49).

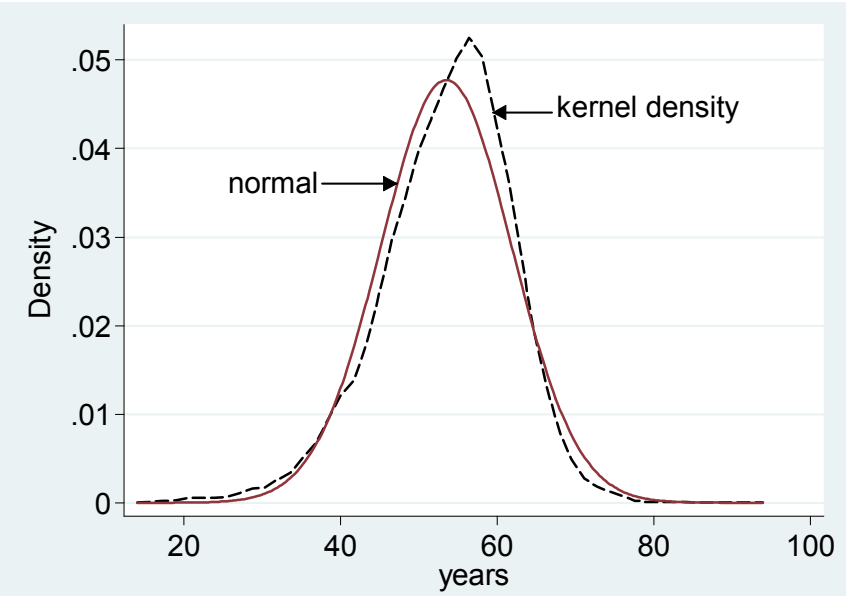


Figure 49. Kernel density plot for age distribution in matched RS and THR female groups

The observed time to revision was far shorter for RS than THR recipients (Figure 50).

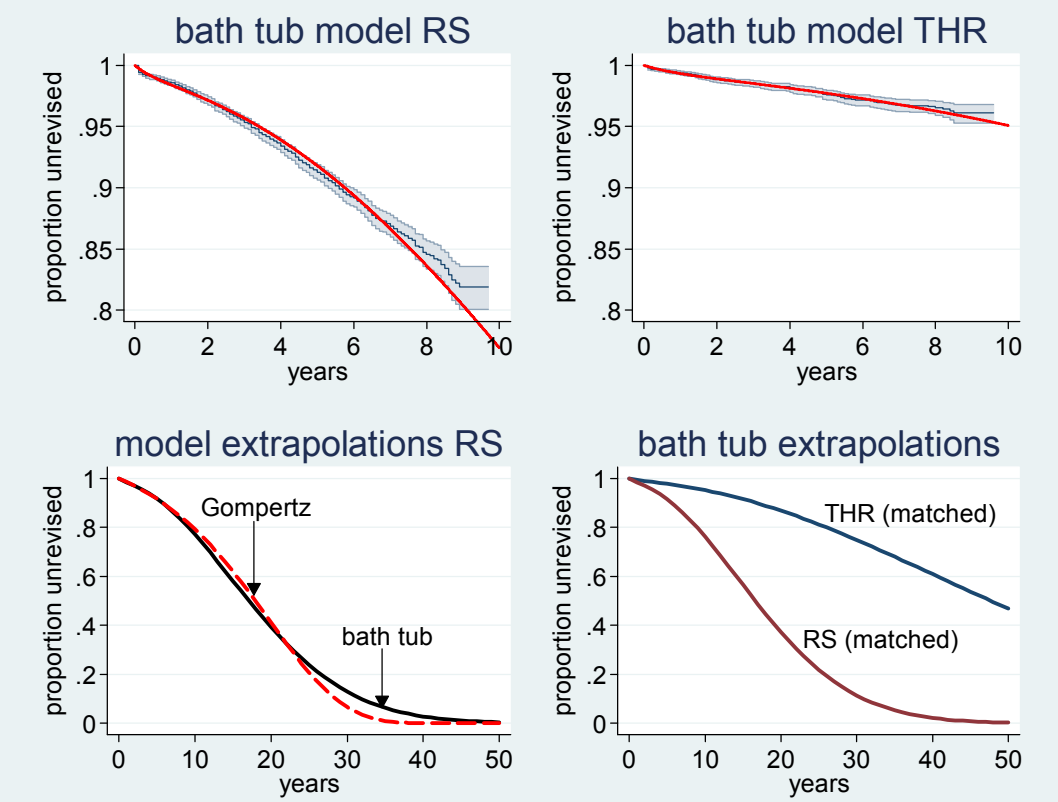


Figure 50. Observed revision (95% CI) & bath tub models for RS & THR

Also shown is Gompertz model (G) for RS.

For RS Gompertz, bathtub and Weibull models provided good fits and each predicted increasing hazard beyond the observed data; according to AIC scores and cumulative hazard plots Gompertz and bath tub were the better fits (Appendices 20 & 21) and predicted similar revision beyond the observed data.

For THR patients the bath tub fit was as good as alternatives (Appendix 18) 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 bath tub, lognormal and Weibull models (Appendices 20 & 21). For the economic analysis the bath tub model was adopted for both RS and THR groups. The predicted requirement for revision at 10, 20 and 30 years using the bath tub model is shown in Table 65.

Table 65. Bath tub modelled percentage of patients requiring revision (females aged 53.5)

Intervention	10 years	20 years	30 years
RS	23.1	61.2	87.6
THR	4.8	13.2	25.2

9.3.5 Comparison of RS and THR according to gender (males)

Each of the matched groups (n = 17,322) had mean age of 57.1 years (SD 8.03; range 16 to 89) and identical age distribution (Figure 51).

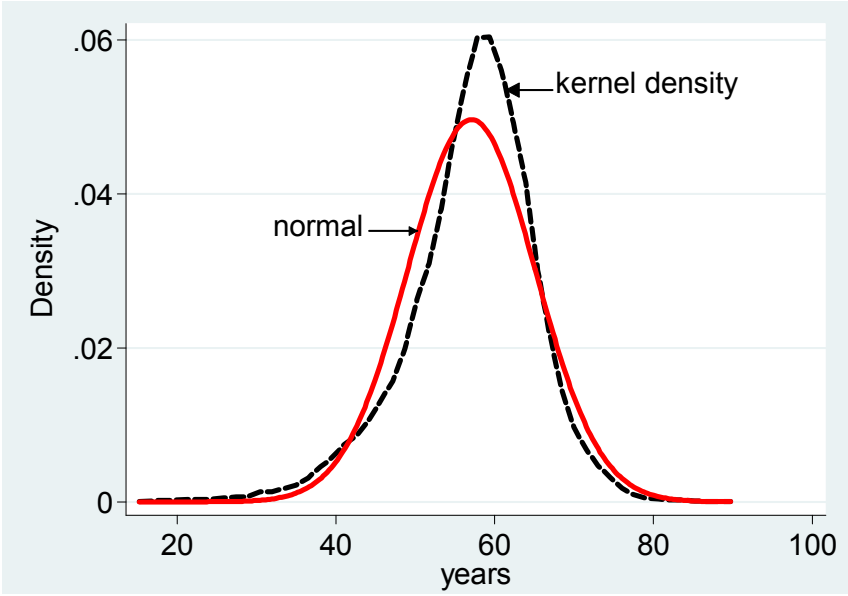


Figure 51. Kernel density plot for age distribution in matched RS and THR male groups

The observed revision rate was higher for RS than THR (Figure 52). Parametric fits are presented in Appendix 18. The bath tub distribution produced the lowest AIC scores and visually the superior fit (Appendices 18 & 20); cumulative hazard plots in Appendix 21. Apart from bath tub, the models predicted decreasing hazard upon extrapolation (Appendix 18). For the economic analysis the bath tub model was adopted for both RS and THR groups. The predicted requirement for revision at 10, 20 and 30 years is shown in Table 66.

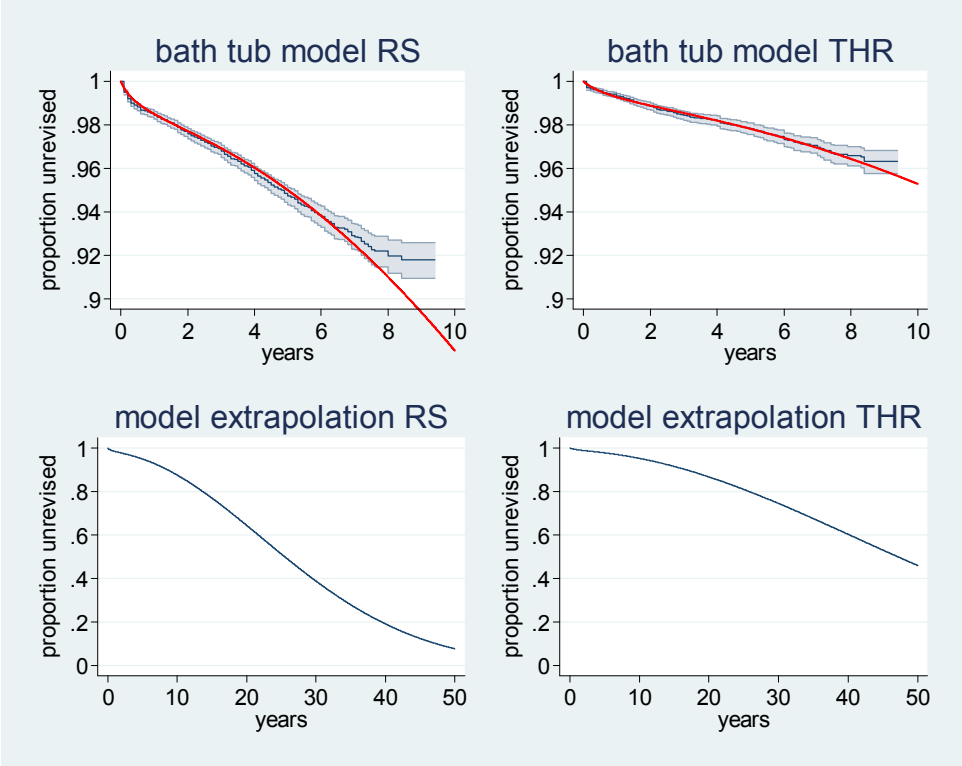


Figure 52. Observed revision (95% CI) and bath tub models for RS & THR males

Table 66. Bath tub modelled percentage of patients requiring revision (males aged 57.1 years)

Intervention	10 years	20 years	30 years
RS	12.4	35.6	61.2
THR	4.7	13.2	25.5

9.3.6 Comparison of THR revision rates according to gender and age; men > 65 years old

Figure 53 shows the observed time to revision for male patients over 65 years of age according to category of THR prosthesis. Revision was less frequent for CePoC (E) than for other categories. Parametric fits to the observed data are shown in Appendix 18, AIC values for models in Appendix 20 and diagnostic plots in Appendix 21. Visually and by AIC scores the bath tub and lognormal models generated best fits except for the CePoC (E) prosthesis for which the bath tub model did not resolve. In view of the advanced age of these patients after accumulating nine years follow up it was considered that an increasing hazard (bath tub) for revision was unlikely and therefore the lognormal model was used for the economic base case. The extrapolations shown in Figure 53 apply for patients aged 70 years.

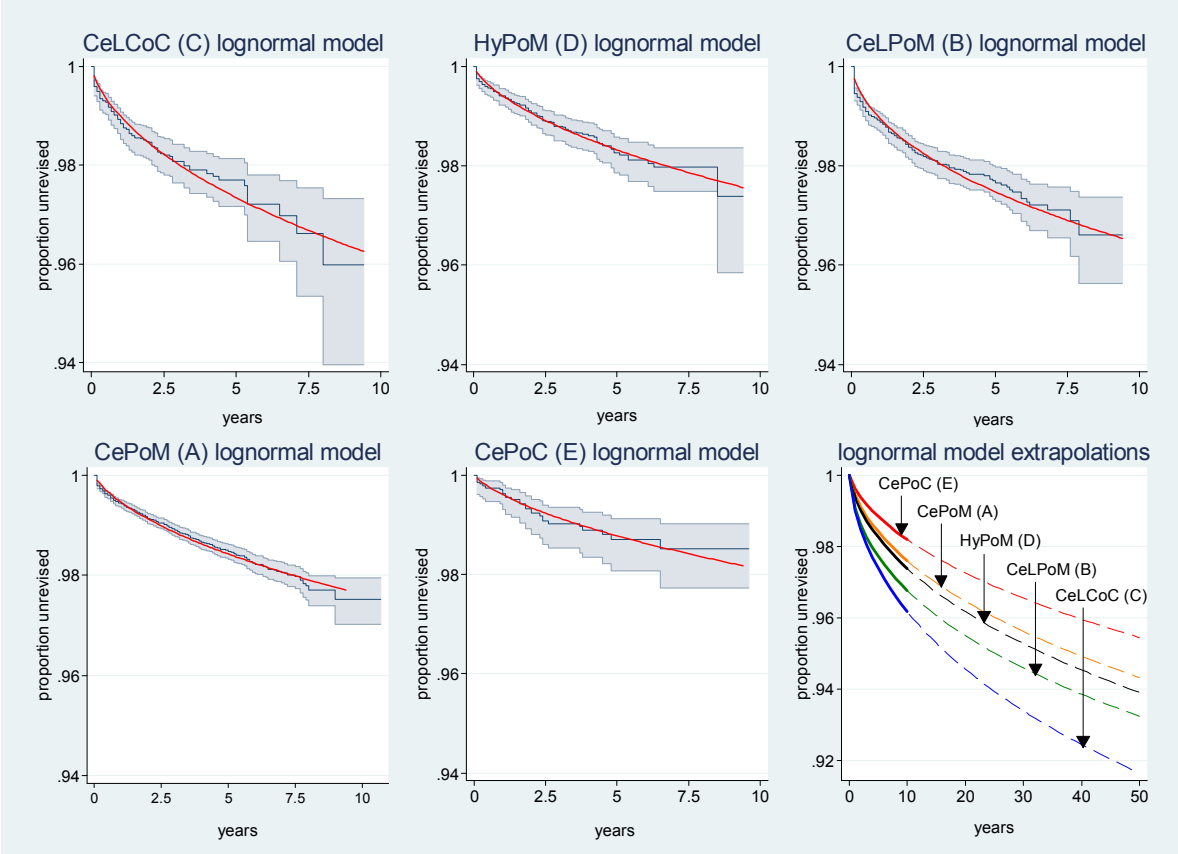


Figure 53. Observed revision (95% CI) for males > 65 years old & lognormal models for THR categories

The model predicted requirements for revision at 10, 20 and 30 years is summarised in Table 67.

Table 67. Lognormal modelled percentage of patients requiring revision (males > 65 years)

THR category	10 years	20 years	30 years
CePoM (A)	2.4	3.5	4.4
CeLPoM (B)	3.6	4.9	5.9
CeLCoC (C)	3.9	5.5	6.7
HyPoM (D)	2.5	3.7	4.6
CePoC (E)	1.9	2.9	3.6
Percentages refer to the mean age of patients in each category			

9.3.7 Comparison of THR revision rates according to gender and age; women > 65 years old

Figure 54 shows the observed time to revision for female patients over 65 years of age according to category of THR prosthesis. Revision was less frequent for CePoC (E) than for other categories. Parametric fits to the observed data are shown in Appendix 18, AIC values for models in Appendix 20 and diagnostic plots in Appendix 21. Visually and by AIC scores the bath tub and lognormal models

generated best fits except for the CePoC (E) prosthesis for which the bath tub model did not resolve. In view of the advanced age of these patients after accumulating nine years follow up it was considered that an increasing hazard (bath tub) for revision is unlikely and therefore the lognormal model was used for the economic base case. The extrapolations shown in Figure 54 apply for patients aged 70 years. Predicted requirement for revision at 10, 20 and 30 years is summarised in Table 68.

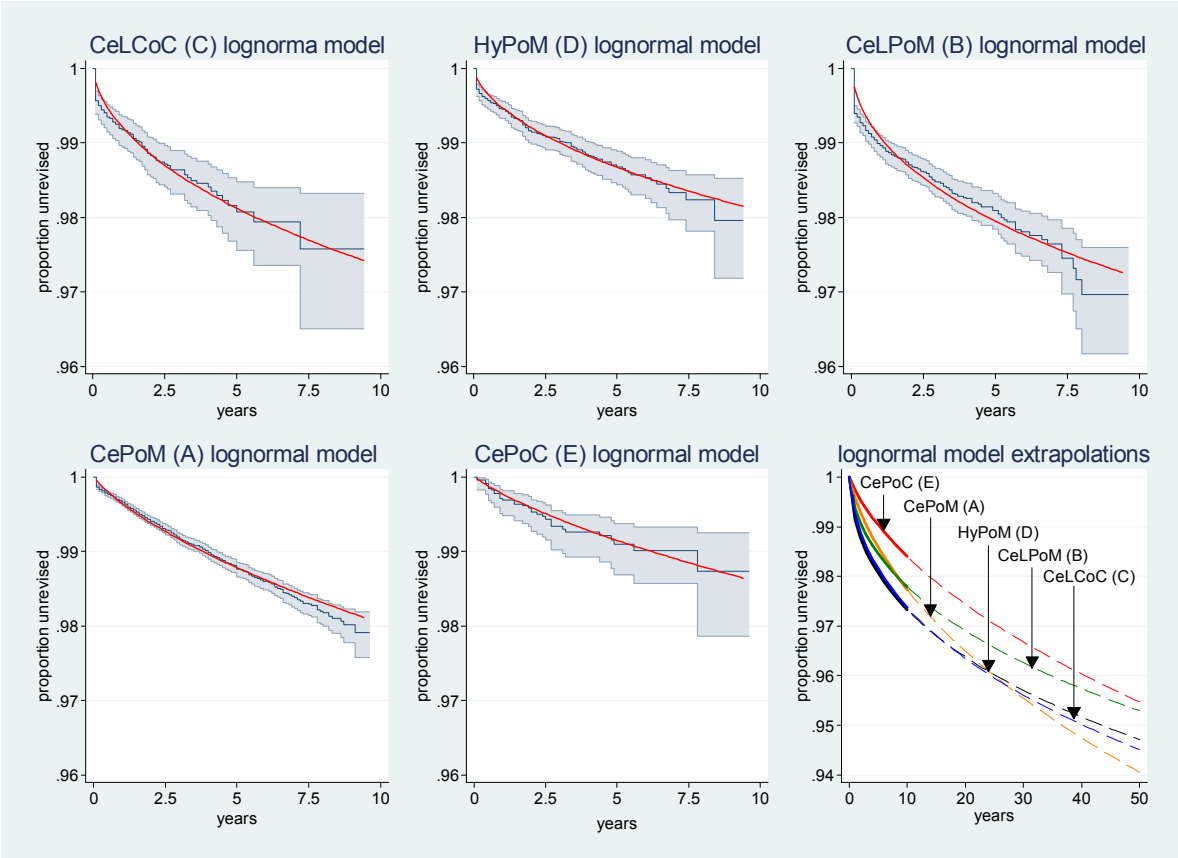


Figure 54. Observed revision (95% CI) for females > 65 years old & lognormal models for THR categories

Table 68. Lognormal modelled percentage of patients requiring revision (females > 65 years)

THR category	10 years	20 years	30 years
CePoM (A)	2.0	3.1	3.9
CeLPoM (B)	2.8	3.8	4.5
CeLCoC (C)	2.7	3.7	4.4
HyPoM (D)	1.9	2.7	3.3
CePoC (E)	1.4	2.3	3.0
Percentages refer to the mean age of patients in each category			

9.3.8 Comparison of THR revision rates according to gender and age; men < 65 years old

Figure 55 shows the observed time to revision for male patients less 65 years of age according to category of THR prosthesis. Parametric fits to the observed data are shown in Appendix 18 and AIC values for models are summarised in Appendix 20. Cumulative hazard plots are shown in Appendix 21. Observed revision was less frequent for CePoC (E) than for other categories. According to AIC values (and visually) the bath tub model provided a superior fit for categories (B, C and D) followed by the lognormal model. For categories A and E there were only trivial differences in AIC values for bath tub and lognormal models. On extrapolation of bath tub models the CePoM category becomes superior to CePoC after about 25 years follow up. Transition probabilities for economic analysis were based on bath tub models (base case for the subgroup) and lognormal models were used in sensitivity analysis. The extrapolations of bath tub models shown in Figure 55 apply for patients aged 50 years.

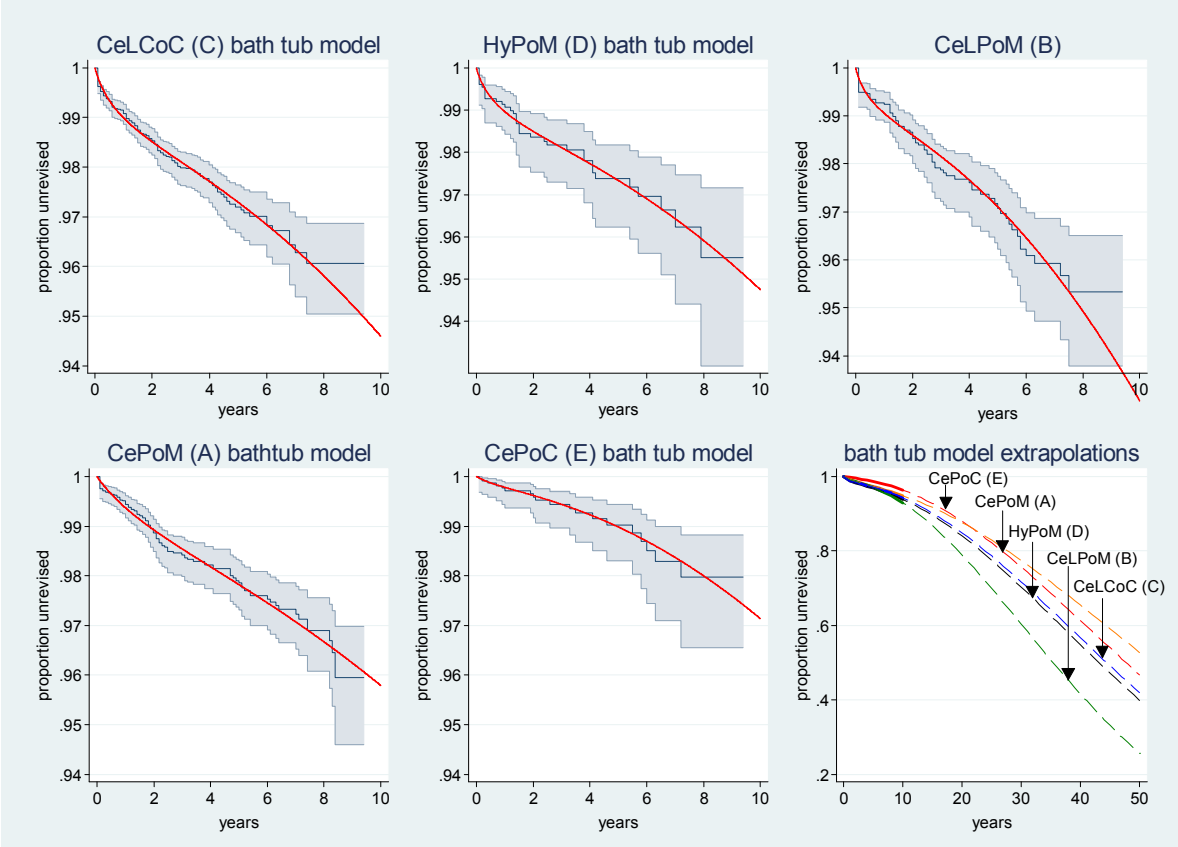


Figure 55. Observed revision (95% CI) for males < 65 years old & bath tub models for THR categories

The bath tub predicted requirement for revision at 10, 20 and 30 years is summarised in

Table 69.

Table 69. Bath tub modelled percentage of patients requiring revision (males < 65 years)

THR category	10 years	20 years	30 years
CePoM (A)	4.2	10.3	18.9
CeLPoM (B)	6.9	20.7	39.0
CeLCoC (C)	5.4	14.3	27.0
HyPoM (D)	5.3	13.8	26.0
CePoC (E)	2.9	8.5	19.7
Percentages refer to the mean age of patients in each category			

9.3.9 Comparison of THR revision rates according to gender and age; women < 65 years old

Table 56 shows the observed time to revision for female patients less 65 years of age according to category of THR prosthesis. Observed revision was less frequent for CePoC (E) than for other categories. Parametric fits to the observed data are shown in Appendix 18 and AIC values for models are summarised in Appendix 20. Cumulative hazard plots are shown in Appendix 21. According to AIC values and visual inspection the bath tub model provided a superior fit to observed data for categories A, C, D and E, but failed to resolve for category B (CeLPoM). Of the tested models for category B, each except for exponential, generated decreasing hazard beyond the observed data. For the economic model the bath tub 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 70.

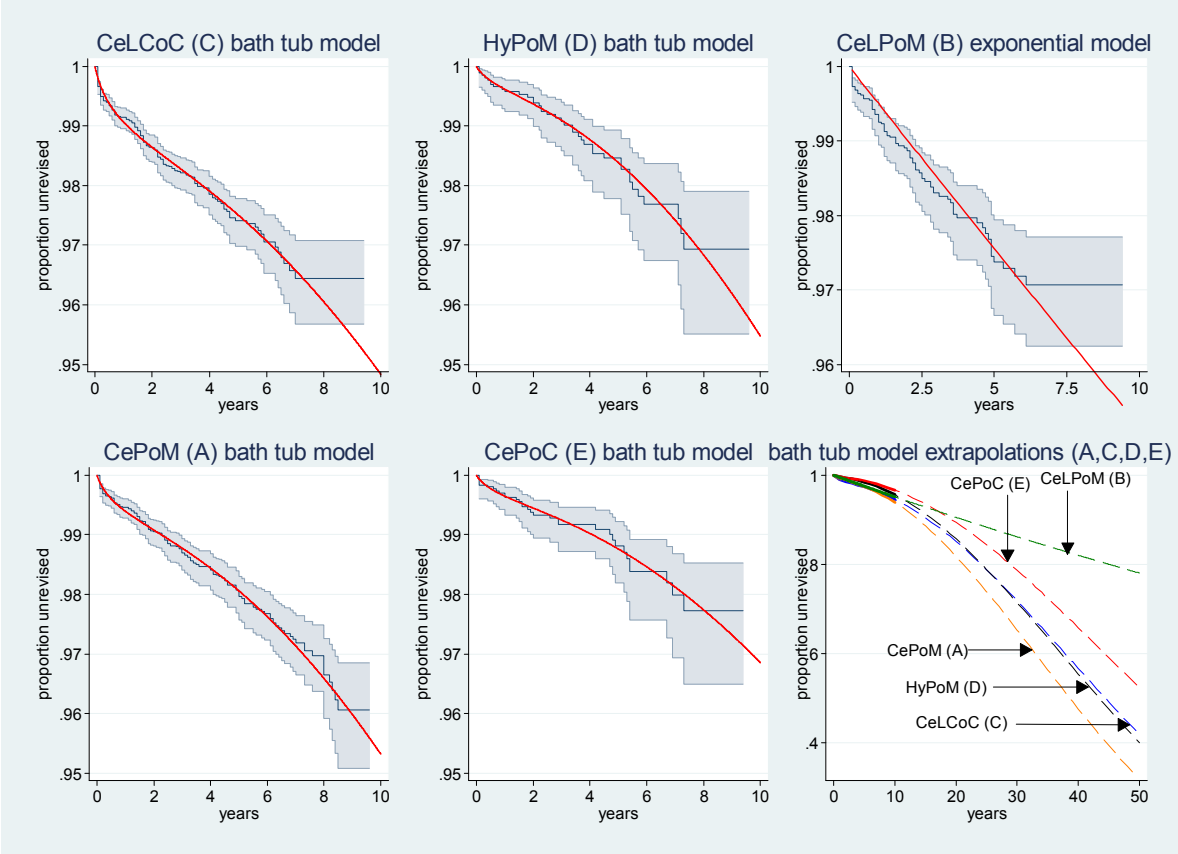


Figure 56. Observed revision (95% CI) for females < 65 years old & bath tub models for THR categories

n.b. a bath tub model did not resolve for category B and so an exponential model was used. The extrapolations of models shown in apply for patients aged 50 years

Table 70. Modelled percentage of patients requiring revision (females < 65 years)

THR category	10 years	20 years	30 years
CePoM (A)	4.7	14.3	28.0
CeLPoM (B)	4.8	9.4	13.8
CeLCoC (C)	5.2	14.2	27.1
HyPoM (D)	4.5	14.9	29.7
CePoC (E)	3.1	10.0	20.3
Percentages refer to the mean age of patients in each category n.b. Bath tub models were used for each category other than B for which an exponential model was used			

9.3.10 Comparison of revision rates with NICE benchmark

The two previous TA guidance documents (TA44 and TA2) gave a 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 derived from an assessment of technologies based on data from approximately 15-20 years ago. Table 71 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 to E, the revision rate at 10 years is within half the benchmark rate, the CePoC (E) prosthesis performing better than the rest. Category A to E THR patients age matched to RS recipients similarly experienced revision rates less than half those of the benchmark, and this also nearly applied for the revision rate observed for all THR patients in the NJR.

In contrast, the revision rates for RS recipients as a whole or for the RS patients after age matching with THR recipients for both genders substantially exceeded the benchmark; the rate for women reached 23.1% and for men 12.4%

Table 71. 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	Category A to E matched to RS (n 26,643)	4.7
THR	All NJR patients (n 386,566)	5.2
THR	All CePoM (A) (n 125,285)	2.8
THR	All CeLPoM (B) (n 37,874)	3.9
THR	All CeLCoC (C) (n 34,754)	4.7
THR	All HyPoM (D) (n 28,471)	3.0
THR	All CePoC (E) (n 12,705)	2.1

These results imply that a new benchmark lower than 10% at 10 years would now appear appropriate for THR technologies, and that RS technologies may require considerable improvement to meet the 10% benchmark.

9.4 Flexible parametric modelling

Several recent analyses of revision rates for patients in the NJR have employed the flexible parametric procedure of Parmar and Lambert.[352] 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 with the purpose of finding if conclusions based on methods described above might be at odds with results from flexible parametric modelling.

In general (see Appendix 22) flexible parametric models generated good fits to the KM estimates of observed revision rates, in some instances AIC scores were as good as or better than 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 Weibull and lognormal models (Appendix 22); as with base case bath tub and with lognormal models the CePoC (E) prosthesis provided the lowest modelled revision rate. With regard to comparison between RS and THR, for both males and females, as with the base case bath tub model, flexible modelling yielded considerably higher rates of revision than did lognormal or Weibull (Appendix 22).

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.

9.5 Discussion, methods of modelling revision rates

In the NJR twice as many men than women received RS, whereas 1.7 times as many women than men received THR categories A to E, while the mean age for RS recipients was nearly 15 years less than that for THR recipients. The number of THR recipients outnumbered RS recipients by about 10 to one. When observed revision rates over about nine years of follow-up were compared between total THR and total RS populations they were found to be about three times higher for RS. The difference was greater for females than males (nearly four-fold and about two-fold, respectively). When the comparisons were made between RS and most frequently used categories of THR these differences were greater.

All THR categories, for both men and women, had far lower revision rates than RS. Because of the age and gender imbalances between RS and THR populations we used propensity matching by age and gender to generate a THR population that would allow an equitable comparison between RS and 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. Revision rates for RS controlled for age were substantially greater than those for THR. This held for both men and women, and when carried through to the economic analysis this translated to higher cost associated with RS than 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 frequency of use of different categories of prosthesis and upon expert clinical opinion. The selection of the five THR categories was conducted pre-hoc and prior to all analyses of revision rates. Just over 239,000 patients in the NJR received one of the five categories of THR prostheses. The observed revision rates were lowest for CePoC (E) and highest for CeLCoC (C) and CeLPoM (B) prostheses. This reflects practice over the last nine to 10 years.

The age and gender distributions varied between categories; when populations were controlled for differences in age and gender, or were stratified by gender and controlled for age, the lower revision rate for the CoPoC (E) category relative to other categories was not diminished. Also when well-fitting models which predicted either increasing or decreasing hazard on extrapolation were used, the superiority of the CoPoC (E) revision rate was again was upheld. There was insufficient information recorded consistently 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 was complete for patients' age and gender at receipt of THR.

For economic modelling we used the revision estimates from KM analysis. This conforms to the practice of previous hip replacement cost effectiveness models found in the literature. McMinn et al. (2012) aptly define the inference in 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".*[308] This was considered appropriate for the structure of the economic model.

To model revision rates we followed NICE DSU recommendations in first exploring exponential, Weibull, Gompertz, lognormal and loglogistic 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 AIC BIC.[356] However, most economic analyses of hip replacement, notably those of Briggs et al. (2004)[38], Higashi et al. (2011)[272] and Pennington et al. (2013),[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 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 bath tub. We therefore also explored bath tub models.

The NJR observation period for both RS and THR patients extended to about nine years. NICE requires a life time economic model so as 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 other than bath tub predicted a decreasing rate of revision (i.e. decreasing hazard), whereas the bath tub models all described increasing revision rate beyond the observed period. Increasing hazard of revision appears reasonable for patients young at primary hip replacement who might expect to live with their prosthesis for 30 or more years.

For older age groups it may be argued that a model predicting increasing hazard for revision is unsuitable since, 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 for NJR patients over 85 years during the observation period was very low and minor relative to attrition due to death. 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.

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 (bath tub) hazard assumption.[38,44,272] This was supported by the goodness of fit to the observed data according to visual inspection, information criteria and plots of log KM-estimated cumulative hazard versus log modelled cumulative hazard.[350] Published analyses with long follow up of patients also supports increasing revision rates beyond ten years from the primary intervention. Previous authors obtained an overall bath tub hazard by a combining Weibull fit for early failures with a Weibull fit for late failures.[38,44,272] We derived the bath tub hazard directly using the STATA package developed by Crowther and Lambert.[351] This had the advantages of parsimony and of not requiring arbitrary decisions about early and late failures. Higashi and Barendregt (2011)[272] used long term follow up studies for the second Weibull fit so as to obtain 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. (2013)[44] employed a piece-wise procedure to generate the "U" shaped hazard, however after extrapolation this predicted that more than 100% 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 that 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 indicated that these are unrelated to RS head sizes so that comparisons of RS and THR according to head size were not undertaken

9.5.1 Summary

The KM estimated rates of revision during approximately nine years of follow up of NJR patients indicated that the probability of revision differed between interventions. RS required considerably higher frequency of revision than THR; this held across both genders. The five categories of THR selected also differed in observed revision rates with CePoC (E) tending to lower rates than alternative categories; again this held generally across age groups and genders.

For all interventions several parametric models generated good fits to the observed data. The differences between models with 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 rates with time (usually, but not always, these were bath tub models) and those predicting a gradual decrease in rates of revision. Data summarised in Section 10.4, from several studies, the Swedish registry, from the RCT of Kim et al. (2011),[126] and from long follow up observational studies tended to support the proposition of increasing hazard, at least for the first decade or so beyond the nine years of NJR data.[93,126,359-363]

On the other hand it is clear that NJR patients who receive THR in old age (e.g. older than 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 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 THR CePoC (E) relative to other categories was maintained across models that differed in the direction of 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 results from an economic model. Therefore, it may be anticipated that over a lifetime the influence of different modelling approaches to extrapolation (increasing hazard for each intervention or alternatively decreasing hazard for each) might not be of great influence on economic outcome for interventions relative to their observed differences.

Our assessment of THR and RS against the benchmark from TA 2 and TA 44 of 10% at 10 years implies that a new benchmark lower than 10% at 10 years would now appear appropriate for THR technologies, but that RS technologies may still require considerable improvement to meet the 10% benchmark.

10 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 Fitzpatrick et al. (1998) which has been adapted for our decision problem and updated with new data.[364]

10.1 Methods

10.1.1 De novo analysis

10.1.1.1 Patients

We used NJR data to investigate revision rates. Detailed information on this is given in Chapters 7 and 9.

We used propensity matching to match by age and gender 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 objective A (see Chapter 7). Furthermore, we performed subgroup analyses in which RS and THR matched populations were stratified by gender and models of time to revision were controlled for age. For objective B, we compared THR categories A to E irrespective of age and gender in the base case. In sensitivity analysis we controlled for age and gender. Population details are listed in Table 57. For sub-group analysis we stratified by age (< 65years and >65 years) and by gender, and the modelled time to revision was controlled for age. The selection of the subgroup > 65 years reflected a population unlikely to be considered suitable only for THR and not suitable for RS (refer to Table 57 for population details and Figure 24).

10.1.2 Model structure

An economic model was developed based on a Markov, multistate model as shown in Figure 57.

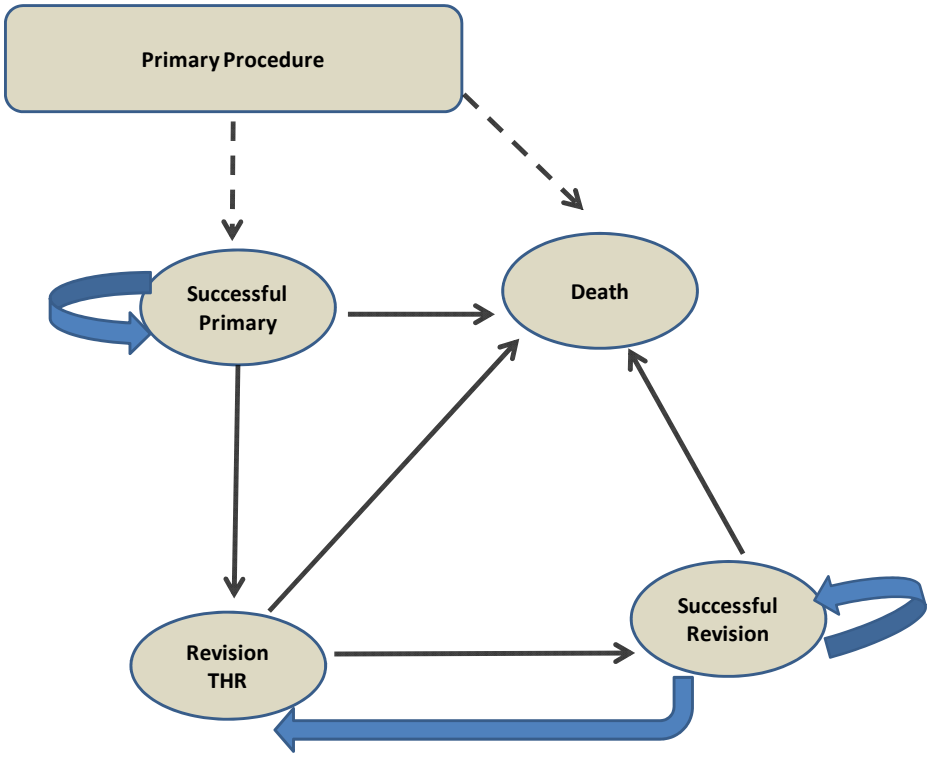


Figure 57. Markov model

based on Fitzpatrick et al. (1998).[364]

In the model, each patient can enter one of four health states following primary surgery:

- Successful primary (RS or THR) surgery (if initial surgery is successful, patients enter this health state)
- 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 only stay in this health state for one cycle
- Successful revision surgery (if revision surgery is successful, patients enter this health state)
- Death (is an absorbing health state and patients may enter this state due to operative mortality or due to death from other causes).

For RS versus THR and for different categories of THR compared with each other (objective B), similar model were built (Figure 57); with different estimates of transition probabilities, utilities and costs.

Cycle length for each model was set at one year and transitions between each health state occur at the end of each cycle. Prior to 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 was agreed upon 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 due to 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 questions, we adopted a 10-year and a lifetime horizon. The 10-year time horizon reflects observed individual patient data from the NJR, and the lifetime horizon follows the recommendation by NICE that the time horizon should be sufficiently extended to capture all benefits likely to accrue from the intervention.[365] The analysis was conducted from the perspective of the NHS and PSS. All costs are in pounds sterling (£) in 2011/2012 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. [365] The key features of the analysis are listed in Table 72.

Table 72. Key features of the analysis

Element of health technology assessment	Reference case	Section in 'Guide to the methods of technology appraisal'
Defining the decision problem	Clinical 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).	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 to 5.2.10
Perspective benefits	All health effects on individuals	5.2.7 to 5.2.10
Type of economic	Cost-effectiveness analysis	5.2.11 and 5.2.12

evaluation		
Synthesis of evidence on outcomes	Based on NJR database	5.3
Measure of health effects	QALYs	5.4
Source of data for measurement of HRQL	Based on PROMS database (Reported directly by patients and carers)	5.4
Source of preference data for valuation of changes in HRQL	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
HRQL, health-related quality of life; NHS, National Health Service; PSS, Personal Social Services; QALY(s), quality-adjusted life year(s)		

10.1.3 Base-case analyses

For the base-case analysis, we estimated the cost-effectiveness of THR against 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 (2013 email from NHS Supply Chain; unreferenced).

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 (2013 email from NHS Supply Chain; unreferenced).

10.1.4 Structural model assumptions

10.1.4.1 Transition probabilities

Time to revision was described according to well-fitting parametric models. (The base case for the comparison of THR versus RS and for the comparison of different THR categories was based on the bathtub model; in sensitivity analysis for THR versus THR a lognormal parametric model was used adjusted for age and gender). The risk of re-revision was based on re-revision rates obtained from the manufacturer's submissions to NICE (sourced from the New Zealand joint registry[366] by the manufacturer).

10.1.4.2 Utilities

Utilities for both models for the base-case analysis were obtained from the PROMS database (see Chapter 8). The mean EQ-5D scores for the successful primary health state and successful revision health state were reduced by the mean EQ-5D scores for the respective age band and gender 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 different types of THR.

10.1.4.3 Costs

For the comparison of THR with RS and for different types of THR, prices of primary prostheses were based on the list prices obtained from the NHS supply chain. We assumed that for the THR versus RS comparison, if initial RS surgery failed, the patient would then be revised with a THR prosthesis and not RS prosthesis. The price of revision prosthesis and re-revision prosthesis were obtained from Vanhegan et al. (2012) based on a weighted average of mean costs of all revision procedures.[291] For the THR versus THR comparison, we assumed that if initial THR surgery failed, the same type of prostheses was used for each category. Hence, we included the mean implant cost from Vanhegan et al. (2012) based on a weighted average of mean costs of all revision procedure.[291]

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 due to 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.[19,364] (see Table 73)

Table 73. Summary of assumptions

Parameter	Assumptions
Transition probabilities	Time to revision was assumed to be described according to well-fitting parametric models. The risk of re-revision was based on re-revision 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 of RS with THR and different types of THR.
Costs	For the comparison of THR with RS and different types of THR, price of primary prosthesis were based on the list prices obtained from the NHS supply chain.

	The price of revision prosthesis and re-revision prosthesis were obtained from Vanhegan et al. (2012) based on a weighted average of mean costs of all revision procedure.[291]

10.1.5 Estimation of model parameters

10.1.5.1 Resource use and cost inputs

Resource use and associated costs were required for the following health states:

1. Cost of successful primary procedure
2. Cost of revision procedure
3. Cost of 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 costs of maintenance.

10.1.5.2 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: Edlin et al. (2012), Vale et al (2002) and Vanhegan et al (2012).[19,40,44 291].

Edlin et al (2012) reported a cost-utility analysis of RS versus THR of a randomised controlled trial using a NHS and PSS perspective and costs were reported as UK £ sterling in 2009/2010 prices. The study used Healthcare Resource Group v4 (HRG4) reference costs combined with NHS Trust finance department list prices for implants and individual patient data on length of stay (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 initial operation/care, subsequent inpatient, outpatient, primary and community care, aids and medications, as well as private and social costs.

Vale et al. (2002) assessed the effectiveness and cost-effectiveness of RS compared to 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 £ sterling in 2000/2001 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 operating theatre, staff, x-rays, outpatient visits and first year follow-up costs were reported.

Vanhegan et al. (2012) investigated the costs of revision THR.[291] Costs were reported in UK £ sterling in 2007/2008 prices and were obtained from the finance department of the tertiary centre and included costs of implant, materials and augmentation, use of the operating theatre and recovery room, the inpatient stay and costs of 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, peri-prosthetic 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 2011/2012 prices by applying the projected health service cost index (HSCI). [367] It is also important to mention that none of the studies identified in the literature included cost per component of prostheses as grouped in our analysis.

10.1.5.3 Base-case cost inputs: RS vs. THR

Cost of primary THR or RS included operation costs, prosthesis costs, hospital ward costs, and follow-up costs.

10.1.5.4 Cost of successful primary procedure (THR or RS) costs

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 (2013 email from NHS Supply Chain; unreferenced). 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 (see Table 74). In real life these prices are often discounted (using a discount de-escalator based on the volume of the purchase).

Table 74. RS prosthesis cost as reported by the NHS supply chain

Component	Average unit cost (£)	Supplier list price (£)		
		Supplier 1 (£)	Supplier 2 (£)	Supplier 3 (£)
Acetabular cup HA coated	1,583	1,690	1,535	1,523
Resurfacing head cemented	1,031	1,140	865	1,089
Mixing Bowl*	31	N/A	N/A	N/A
Cement (1 pack)*	27	N/A	N/A	N/A
Total cost	**2,672**			

*The price is sourced from one supplier

The cost of the THR prosthesis 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 our categories (A to E) n the RS vs. THR comparison (Table 75).

Table 75. Cost of THR prosthesis

Category	Number of male patients	Number female patients	Total number of patients	Mean cost (£)	Weighted cost (£)
A	6,080	3,812	9,892	1,557	589
B	2,177	741	2,918	3,016	336
C	5,803	2,414	8,217	3,869	1,215
D	1,104	477	1,581	2,650	160
E	2,100	1,459	3,559	1,996	271
Weighted cost of THR prostheses					£2,571

The cost of the surgery itself was assumed to be the same for both THR and RS. The cost of theatre overheads, theatre staff and number of x-rays etc. were taken from Vale et al. (2002)[19] and costs were updated to current prices. [367] The total cost of surgery was estimated at £2,805 (see Table 76).

Table 76. Total cost of surgery

Resource use	1996 prices		2011/2012 prices
	Primary THR (units)	**Total cost (£)**	**Total cost (£)**
Theatre overheads	134	655	1,799
Theatre staff	-	232	637
Number of x-rays	6	134	368
Total cost per patient			£2,805

The average length of stay was based on point estimates as reported in Edlin et al. (2012).[40] The total cost of inpatient stay for RS was estimated to be £1,628. This was based on the average cost per day of hospital stay at £296, multiplied by the average length of stay at 5.5 days, as reported by Edlin et al. (2012).[40] The average length of stay for THR was 5.7 days and the total cost of inpatient stay for THR was estimated to be £1,687. RS was associated with a slightly shorter length of stay (5.7 vs. 5.5 days); although this difference was not statistically significant, we assigned this slightly shorter length of stay so as not to overestimate costs of RS.

10.1.5.5 Cost of revision procedure (THR or RS)

Costs for revision were assumed to be the same for both the THR and the RS groups. The cost of a revision hip arthroplasty was obtained from Vanhegan et al. (2012); [291] the data were based on 305 successive revisions following THR in 286 patients between 1999 and January 2008. In the study, patient specific resource use data were reported for implant cost, materials, theatre cost, use of

recovery room, inpatient stay, physiotherapy, occupational therapy, pharmacy, radiology and laboratory costs, with costs based on NHS 2007/2008 rates for payment by results (PbR).

Costs were inflated to 2011/2012 prices by applying the projected health service cost index (HSCI).[367] Importantly, the study also reported mean costs for revision surgery in aseptic cases, septic revision, peri-prosthetic fracture and for dislocation. Hence, the cost of revision was calculated based on a weighted average of mean costs of all revision procedures (see Table 77).

Table 77. Cost of revision

Indication	Number of patients	Mean cost (£) (2007/2008 prices)	Mean cost (£) (2011/2012 prices)
Aseptic loosening	194	11,897	13,226
Deep infection	76	21,937	24,387
Peri-prosthetic fracture	24	18,185	20,216
Dislocation	11	10,893	12,109
Weighted average			£16,517

10.1.5.6 Cost of successful revision procedure (THR or RS)

The cost of follow-up post primary THR or RS was obtained from Edlin et al. (2012),[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 and the RS arms. The NHS and social care costs of follow-up were £394 for the THR arm and £501 for the RS arm at 12 months (see Table 78).

Table 78. Cost of successful revision procedure (THR/RS)

Costs	2009/2010 prices (£)		2011/2012 prices (£)	
	Total cost RS	**Total cost THR**	**Total cost RS**	**Total cost THR**
Outpatient	360	276	383	294
Primary/Community	63	49	67	52
Aids & adaptations	21	21	22	22
Medications	27	24	29	26
Total cost			501	394

10.1.5.7 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 THR vs. RS comparison. The cost of primary THR included 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 inpatient stay was estimated to be £1,687, based on the average cost per day of hospital stay, multiplied by the average of length of stay (5.7 days), as reported in Edlin et al. (2012).[40] The total cost of surgery including x-ray, theatre time, staff and overheads was estimated at £2,805. [367] Outpatient costs and other follow-up costs were estimated to be £394 based on Edlin et al. (2012)[40] (see Table 78).

10.1.5.8 Prosthesis cost

We were not able to use published costs for the costs of the prosthesis because prostheses were grouped as cemented, cementless or hybrid rather than based on the separately identifiable prosthesis components as categorised in our analysis (Categories A-E). Our base-case cost per category of prosthesis was obtained from the NHS supply chain (2013 email from NHS Supply Chain; unreferenced). Anonymised information was available detailing list price per component for all five categories. The cost data from the commonest five suppliers using their most common brands of implant were available and an average cost was calculated. Again this is subject to a volume deescalator in price for the NHS (see Table 79).

The pricing of a bone cement pack including bone cement, mixing devices and pressuriser was only available from one supplier. We have itemised the cost of a bone cement pack for cemented stem and cup, and cemented stem alone (see Table 80).

Table 79. Prosthesis cost

Component	Average unit cost (£)	Supplier 1	Supplier 2	Supplier 3	Supplier 4	Supplier 5
A - CePoM						
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	N/A	19	76	N/A	N/A
Bone cement plug	58.38	44.5	49	N/A	81	59
Cemented stem & cup extras	203.10					
Total	**£1,557.38**					
B - CeLPoM						
Cementless HAC stem	1,342.20	1,370	1,129	1,110	1,816	1,286
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	**3,015.60**					
C - CeLCoC						
Cementless HAC stem	1,342.20	1,370	1,129	1,110	1,816	1,286
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	**3,868.80**					
D - HyPoM						
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	N/A	19	76	N/A	N/A
Bone cement plug	58.38	44.5	49	N/A	81	59
Fixation screw	85.60	82	96	73	74	103
Cemented stem extras	163.90					
Total	**2,649.78**					
E - CePoC						
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	N/A	19	76	N/A	N/A
Bone cement plug	58.38	44.5	49	N/A	81	59
Cemented stem & cup extras	203.10					
Total	**£1,995.98**					

Table 80. Cost of bone cement pack

Pack	Component	Total cost (£)
Cemented stem and cup	Cement 40 gram & 80 gram pack	203.10
	Cement syringe	
	Femoral pressuriser	
	Cement mixing pot	
	Acetabular pressuriser	
Cemented stem	Cement 80 gram	163.90
	Cement syringe	
	Femoral pressuriser	
	Cement mixing pot	

10.1.6 A summary of the transition probabilities, utilities and cost inputs to the cost-utility model

The justification for 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 from the NHS Supply Chain since alternative sources of information were lacking.

Utilities were calculated from information in the PROMs database. This was justified because it represented patent centred EQ-5D 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 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 81.

Table 81. Bathtub parameters for comparison RS vs. THR and THR vs. THR

Comparison	Prosthesis	BT alpha	BT beta	BT 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	CePoM (A)	0.0003396	0.0083374	2.163733
Base case	CeLPoM (B)	0.0004045	0.0337383	6.832735
Base case	CeLCoC (C)	0.0005333	0.0236369	4.051712
Base case	HyPoM (D)	0.0003642	0.0158328	4.68618
Base case	CePoC (E)	0.0001935	0.0039017	0.6967542

Table 82 shows a summary of the inputs (transition probabilities, utilities and costs) used in the base-case analysis.

Table 82. Summary of transition probabilities, utilities and cost inputs for base-case analysis

Transition probabilities					
Health state	**Mean value**	**SE**			**Source**
Surgical mortality*	0.0050	0.001			NJR annual report[48]
Risk of re-revision	0.0326	N/A			DePuy submission
Utility inputs					
Utilities	**Mean value**	**SE**	**Beta distribution Parameter α**	**Beta distribution Parameter β**	**Source**
Age 50-60	0.7529	0.004	1,296	488	PROMS[368]
Age 60-70	0.7789	0.002	7,397	2,427	PROMS[368]
Age 70-80	0.7637	0.002	22,244	6,315	PROMS[368]
Age 80+	0.7210	0.003	28,054	8,681	PROMS[368]
Revision surgery	0.5624	0.340	9,092	3,518	PROMS[368]
Cost inputs					
RS versus THR					
Health state	**Mean value £**	**SE**	**Gamma distribution Parameter α**	**Gamma distribution Parameter β**	**Source**
RS comparison					
Prosthesis cost	£2,778	N/A	N/A	N/A	NHS Supply Chain
Surgery costs (excluding prosthesis)	1,485	N/A	N/A	N/A	Vale et al. (2002) [19]
Hospital inpatient stay	1,628	N/A	N/A	N/A	Edlin et al. 2012[40]
Successful primary RS	501	44	130	4	Edlin et al. (2012) [40]
Revision surgery	16,517	456	1314	13	Vanhegan et al. (2012) [291]
Successful revision surgery	394	30	169	2	Edlin et al. 2012 [40]
THR comparison					
Prosthesis cost	£2,571	N/A	N/A	N/A	NHS Supply Chain

Surgery costs (excluding prosthesis)	1,485	N/A	N/A	N/A	Vale et al. (2002) [19]
Hospital inpatient stay	1,687	N/A	N/A	N/A	Edlin et al. 2012[40]
Successful primary THR	394	30	169	2	Edlin et al. (2012) [40]
Revision surgery	16,517	456	1314	13	Vanhegan et al. (2012) [291]
Successful revision surgery	394	30	169	2	Edlin et al. 2012 [40]
Different types of THR					
Prosthesis cost	**Mean value £**	**SE**	**Gamma distribution Parameter α**	**Gamma distribution Parameter β**	**Source**
A – CePoM	1,557	N/A	N/A	N/A	NHS Supply Chain
B - CeLPoM	3,017	N/A	N/A	N/A	NHS Supply Chain
C - CeLCoC	3,869	N/A	N/A	N/A	NHS Supply Chain
D - HyPoM	2,650	N/A	N/A	N/A	NHS Supply Chain
E - CePoC	1,996	N/A	N/A	N/A	NHS Supply Chain
Other costs	**Mean value £**	**SE**	**Gamma distribution Parameter α**	**Gamma distribution Parameter β**	**Source**
Surgery costs (excluding prosthesis)	1,485	N/A	N/A	N/A	Vale et al. (2002) [19]
Hospital inpatient stay	1,687	N/A	N/A	N/A	Edlin et al. 2012[40]
Successful primary THR	394	30	169	2	Edlin et al. (2012) [40]
Revision surgery	16,517	456	1314	13	Vanhegan et al. (2012) [291]
Successful revision surgery	394	30	169	2	Edlin et al. 2012 [40]

*surgical mortality was the same for THR, RS and revision

10.1.7 Cost-effectiveness analysis

The base-case analysis is based on costs and outcomes for all THR and RS patients over two time horizons (10-year and lifetime).

For the RS vs. 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 to E, and transition probabilities were calculated using bathtub model fits (predicting increasing hazard beyond the 10-year observation period). For the base case analysis THR vs. THR, transition probabilities were calculated using bathtub model fits for categories A to E.

We report total mean costs and total mean QALYs related to THR and RS, and incremental costs per QALY (ICER) 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 which were more costly and less effective, these were eliminated (extended dominance). With the remaining options, we calculated incremental costs per QALY gained.

We present firstly 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 1,000 simulations. The results from these simulations were plotted on a cost-effectiveness plane with 95% confidence intervals. Each point is a simulation from the probabilistic analysis. The plot illustrates the uncertainty surrounding incremental costs and QALYs for the two groups being compared. We also produced cost-effectiveness acceptability curves (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 willingness to pay threshold from £0 to £50,000. The perspective is from the UK NHS and PSS. Discounting of costs and benefits at 3.5% was undertaken according to UK guidelines. [365]

10.1.8 Sensitivity analyses

Sensitivity analyses were conducted by altering base-case inputs to the model. Several types of sub-group and scenario analyses were explored encompassing changes to the RS versus THR and the THR versus THR comparisons.

10.1.8.1 Sub-group analyses for RS versus THR and THR versus THR

a. Revision rates were much higher for women receiving RS than for men, and because revision rates vary according to the age of patients, subgroup analyses focussed on comparing populations stratified by gender and controlled for age. Therefore, in the sensitivity analysis we separately compared the cost-effectiveness of RS versus THR for male and females aged 40, 50, and 60 years at the time of primary implant, using age-matched population and a bathtub model stratified by gender and controlled for age.

b. For THR versus THR, the modelled time to revision was stratified by age (less than 65 years and more than 65 years) and gender, and models were controlled for age. We undertook these subgroup analyses because the use of different categories of THR prostheses differed by age and gender, and since recipients of hip replacement interventions aged over 65 years approximate a population unlikely to be considered candidates for RS. We compared the cost-effectiveness of THR vs. THR for patients aged under 65 years (40, 50 and 60 years) using a bathtub model and for patients aged over 65 years (70 and 80 years) using a lognormal model (see Table 83).

For sub group analyses, mean EQ-5D index scores were split by gender and age band (see Table 84).

Table 83. Sub-group analysis – Time to revision for RS vs. THR and THR vs. THR

Gender	Prosthesis	BT alpha	BT beta	BT gamma	BT age coefficient
RS vs. THR (matched)- Bathtub parameters					
Males	RS	0.0020179	0.0370237	4.443342	-0.0380901
Males	THR	0.0006006	0.0135972	2.384484	-0.0258836
Females	RS	0.0044984	0.0280047	2.558539	-0.0118076
Females	THR	0.0005964	0.0099966	1.314233	-0.016463
THR vs. THR – Bathtub parameters					
male < 65 years	CePoM (A)	0.0003869	0.008084	0.7177154	-0.0207576
male < 65 years	CeLPoM (B)	0.0010417	0.0245433	4.822729	-0.0024683
male < 65 years	CeLCoC (C)	0.0006243	0.0212657	3.032461	-0.0110798
male < 65 years	HyPoM (D)	0.0005998	0.0237569	3.576745	-0.0172004
male < 65 years	CePoC (E)	0.0004695	0.0033726	1.782609	-0.0327686
female < 65 years	CePoM (A)	0.0006692	0.0132853	3.675229	-0.0293667
female < 65 years	CeLPoM (B)	not resolved			
female < 65 years	CeLCoC (C)	0.0006154	0.0215004	3.952961	-0.0088734
female < 65 years	HyPoM (D)	0.00076	0.0077105	3.21092	0.0048101
female < 65 years	CePoC (E)	0.0004703	0.0071811	3.211915	-0.0078225
THR vs. THR – Lognormal parameters					
Gender	Prosthesis	LN mu	LN sigma	LN age coefficient	
male > 65 years	CePoM (A)	10.37363	4.075863	0.0020929	
male > 65 years	CeLPoM (B)	10.52551	4.554688	-0.0483328	
male > 65 years	CeLCoC (C)	9.611438	4.12394	-0.0448092	
male > 65 years	HyPoM (D)	10.31021	4.093764	0.0126215	
male > 65 years	CePoC (E)	10.54446	3.971899	-0.0407056	
female > 65 years	CePoM (A)	9.815575	3.636813	0.033098	
female > 65 years	CeLPoM (B)	12.10535	5.138115	-0.0241371	
female > 65 years	CeLCoC (C)	11.471	4.744101	-0.0287428	
female > 65 years	HyPoM (D)	12.18021	4.757849	0.0504173	
female > 65 years	CePoC (E)	10.13035	3.562737	0.0631827	

Table 84. Summary of utilities inputs for sub-group analysis

Utility inputs					
Age group	**Mean value**	**SE**	**Beta distribution Parameter α**	**Beta distribution Parameter β**	**Source**
Males					
40-50	0.736	0.0179	443	159	PROMS[368]
50-60	0.767	0.0066	3133	952	PROMS[368]
60-70	0.762	0.0038	9112	2393	PROMS[368]
70-80	0.790	0.0034	11488	3054	PROMS[368]
80+		0.0071	2816	964	PROMS[368]
Females					
40-50	0.720	0.0129	872	339	PROMS[368]
50-60	0.742	0.0058	4287	1491	PROMS[368]

60-70	0.769	0.0032	13128	3944	PROMS[368]
70-80	0.747	0.0029	16732	5667	PROMS[368]
80+	0.710	0.0048	6305	2575	PROMS[368]
Revision surgery					
Males	0.575	0.009	1496	1106	PROMS[368]
Females	0.553	0.007	2201	1779	PROMS[368]

10.1.8.2 Sensitivity analyses around base-case time to revision for RS versus THR

a. Time to revision: the bathtub model was controlled for age and gender. This was done because the age distributions of the matched populations were somewhat removed from normal distribution (see Figure 36 Chapter 9). Transition probabilities were then calculated for the average population (35% female, aged 55.8 years) (see Table 85).

10.1.8.3 Sensitivity analyses for base case time to revision for THR versus THR

a. Time to revision: the bathtub model was controlled for age and gender. This was done because both age and gender differed between categories and both variables were influenced in the time to revision (see Table 63 Chapter 9). Transition probabilities were then calculated for the age and gender mix across all five categories (63.5% females, aged 71.6 years).

b. Time to revision: a lognormal model was used. This was undertaken because the information criteria 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. Time to revision: a lognormal model controlled for age and gender was used. This was done because both age and gender differed between categories, and both variables were associated with time to revision. Transition probabilities were then calculated for the age and gender mix across all five categories (63.5% females, aged 71.6 years) (see Table 86).

Table 85. Sensitivity analysis – Time to revision for RS vs. THR

Comparison	Prosthesis	BT alpha	BT beta	BT gamma	BT age coefficient	BT gender 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 86. Sensitivity analysis – Time to revision for THR vs. THR

Comparison	Prosthesis	BT alpha	BT beta	BT gamma	BT age coefficient	BT gender coefficient
a. THR vs. THR (bathtub model controlled for age and gender)						
Sensitivity analysis	CePoM (A)	0.0003132	0.008041	2.081738	-0.0236324	0.2120103
Sensitivity analysis	CeLPoM (B)	0.0003712	0.030807	6.827069	0.0014804	0.2144175
Sensitivity analysis	CeLCoC (C)	0.0004542	0.0203098	4.028858	-0.0070475	0.1657326
Sensitivity analysis	HyPoM (D)	0.000317	0.0145044	4.595129	-0.019714	0.2461955
Sensitivity analysis	CePoC (E)	0.0001675	0.0034053	0.680878	-0.0149548	0.1011695
b. THR vs. THR (lognormal model)						
Comparison	Prosthesis		LN mu		LN sigma	
Sensitivity analysis	CePoM (A)		9.738756		3.716562	
Sensitivity analysis	CeLPoM (B)		10.71464		4.573634	
Sensitivity analysis	CeLCoC (C)		9.526446		4.034555	
Sensitivity analysis	HyPoM (D)		10.66382		4.215337	
Sensitivity analysis	CePoC (E)		9.574467		3.481879	
c. THR vs. THR (lognormal model controlled for age and gender)						
Comparison	Prosthesis	LN mu	LN sigma	LN age coefficient	LN gender coefficient	
Sensitivity analysis	CePoM (A)	9.825973	3.730391	0.03258	-0.3417841	
Sensitivity analysis	CeLPoM (B)	10.84608	4.563342	-0.0077298	-0.3729022	
Sensitivity analysis	CeLCoC (C)	9.747396	4.036228	0.0093327	-0.2627816	
Sensitivity analysis	HyPoM (D)	10.85018	4.238437	0.0314349	-0.3886501	
Sensitivity analysis	CePoC (E)	9.729236	3.482196	0.01658	-0.1431533	

10.1.8.4 Sensitivity analyses for cost inputs

For this sensitivity analysis, we varied 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 (see Table 87)
   b. Lowest list price for both RS and THR prostheses (see Table 87)

ii. THR vs. THR comparison
   a. Highest list price for all THR prostheses (see Table 88)
   b. Lowest list price for all THR prostheses (see Table 88)

iii. We assumed a 20% price de-escalator to reflect what the 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 cost of each prosthesis in different categories (see Table 87).

Table 87. Highest and lowest list price for both RS and THR (weighted average of all categories) prosthesis

Comparison	**Base case list price (£)**	**Highest list price (£)**	**Lowest list price (£)**
THR	2,571	3,073	2,180
RS	2,778	2,994	2,487

Table 88. Prosthesis cost for sensitivity analysis – highest and lowest unit costs and using a 20% reduction in prosthesis list price across all categories

Category	**Component**	**Highest average unit cost (£)**	**Lowest average unit cost (£)**	**20% reduction in prosthesis list price: average unit cost (£)**
A - CePoM	Cemented stem	1,789	1,241	1,246
	Metal head			
	Polyethylene cup – cemented			
	Cemented stem centraliser			
	Bone cement plug			
	Cemented stem & cup extras			
B - CeLPoM	Cementless HAC stem	3,774	2,662	2,413
	Metal stem			
	Metal cup – cementless HA			
	Liner- polyethylene			
	Fixation screw			
C - CeLCoC	Cementless HAC stem	4,734	3,507	3,095
	Ceramic head			
	Metal cup – cementless HA			
	Liner ceramic			
	Fixation screw			
D - HyPoM	Cemented stem	2,980	2,219	2,120
	Metal head			
	Metal cup - cementless HA			
	Liner polyethylene			
	Cemented stem centraliser			
	Bone cement plug			
	Fixation screw			
	Cemented stem extras			
E - CePoC	Cemented stem	2,271	1,657	1,597
	Ceramic head			
	Polyethylene cup – cemented			
	Cemented stem centraliser			
	Bone cement plug			
	Cemented stem & cup extras			

10.1.8.5 Sensitivity analyses for utility inputs

In the base case, utility values were obtained from the PROMS dataset. For this sensitivity analysis, utility values were taken from Rolfson et al, 2011.[297] This study reported one year post-operative utility values for 32,396 patients from the Swedish Hip arthroplasty register using a UK EQ-5D tariff. Utility values from the PROMS dataset was applied to re-revision health as in the base-case. The impact of this assumption was only tested for the THR vs. THR comparison and not for the RS vs. THR comparison.

Table 89. Summary of utilities inputs for sensitivity analysis

Utility inputs					
Age group	**Mean value**	**SE**	**Beta distribution Parameter α**	**Beta distribution Parameter β**	**Source**
50-60	0.77	0.0036	10006	2989	Rolfson et al [297]
60-70	0.80	0.0021	28270	7067	Rolfson et al [297]
70-80	0.78	0.0021	30273	8538	Rolfson et al [297]
80+	0.73	0.0035	11350	4197	Rolfson et al [297]

10.1.8.6 One way Sensitivity analysis for Category E vs. Category A (Tornado diagram)

A one way sensitivity analysis was conducted to examine the individual impact of the net monetary benefit of Category E (CePoC) vs. Category A (CePoM). All parameters were varied around the base-case value within the plausible ranges as specified.

10.1.8.7 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 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 vs. THR and THR vs. RS comparisons trials were often based on selective populations or interventions. Studies and provided data on revision rates which were inconclusive with often wide confidence intervals.

10.2 Results of cost-effectiveness analysis

We present here cost-effectiveness deterministic and probabilistic results for RS versus THR and for comparison of different types of THR.

10.2.1 The cost-effectiveness analysis results: RS vs. THR

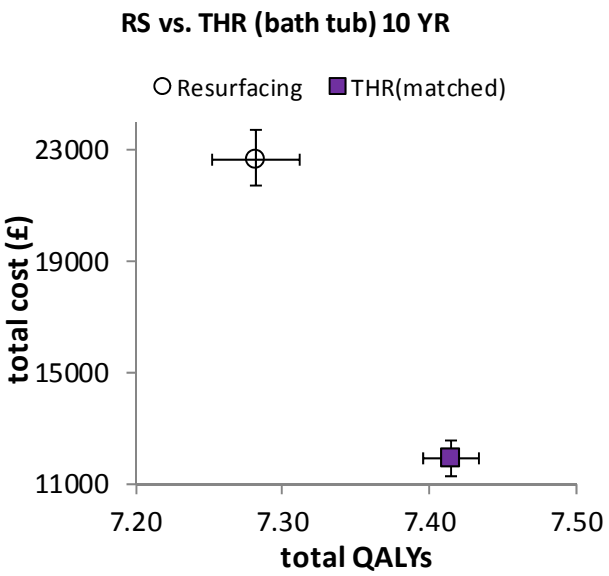
For the base-case analysis, we compared the cost-effectiveness of different types of primary THR compared with RS for people in whom both procedures are suitable.

Table 90 shows the deterministic and probabilistic results for the 10-year and lifetime horizons. For all scenarios, the mean costs for RS were higher than THR; and the mean QALYs were lower. For all scenarios, the incremental cost-effectiveness ratio for RS was dominated by THR; that is, THR was cheaper and more effective than RS.

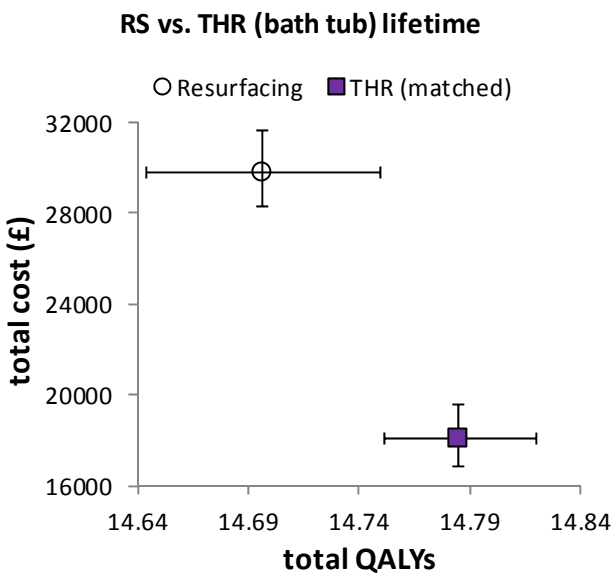
Table 90. Base-case deterministic and probabilistic results for all patients using bathtub model

	RS	THR
Deterministic - 10-year time horizon		
Total mean costs £	22,519	11,879
Total mean QALYs	7.2830	7.4147
Incremental cost £	10,641	
Incremental QALYs	-0.1317	
ICERs (£/QALY)	Dominated	
Deterministic - Life-time horizon		
Total mean costs £	29,603	18,113
Total mean QALYs	14.6968	14.7846
Incremental cost £	11,490	
Incremental QALYs	-0.0879	
ICERs (£/QALY)	Dominated	
Probabilistic - 10-year time horizon		
Total mean costs £	22,615	11,887
Total mean QALYs	7.2823	7.4150
Incremental cost £	10,729	
Incremental QALYs	-0.1327	
ICERs (£/QALY)	Dominated	
Probabilistic - Lifetime horizon		
Total mean costs £	29,770	18,120
Total mean QALYs	14.6963	14.7848
Incremental cost £	11,650	
Incremental QALYs	-0.0885	
ICERs (£/QALY)	Dominated	

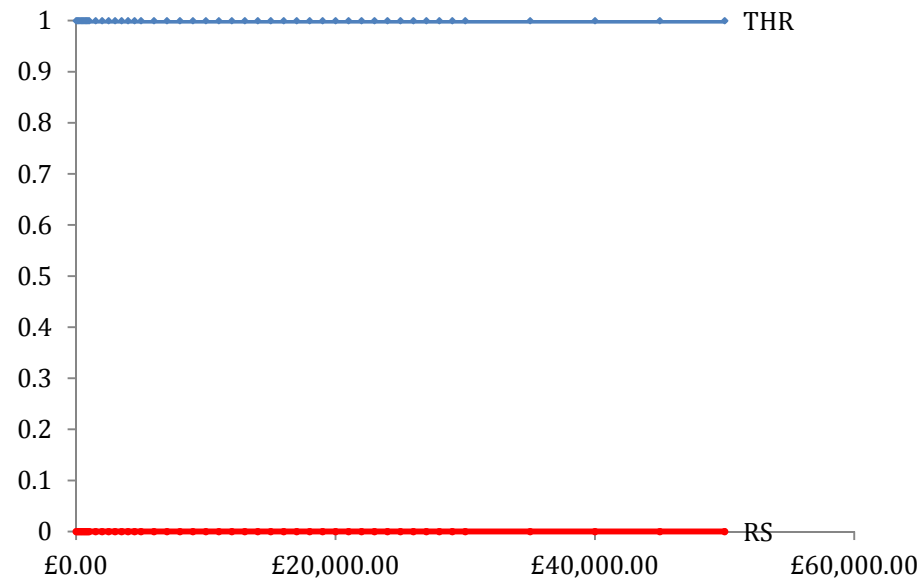
Figure 58 a and b shows the cost-effectiveness planes for THR vs. RS for the 10-year and lifetime horizons. The graph clearly shows that THR dominates RS, as the iterations fall in the north-west quadrant of the plane, that is, RS is clearly more costly and less effective than THR. Figure 58 c and d shows the cost-effectiveness acceptability curves for the two time horizons. For a willingness to pay threshold from £0 to £50,000 per QALY, THR is the more cost-effective option.



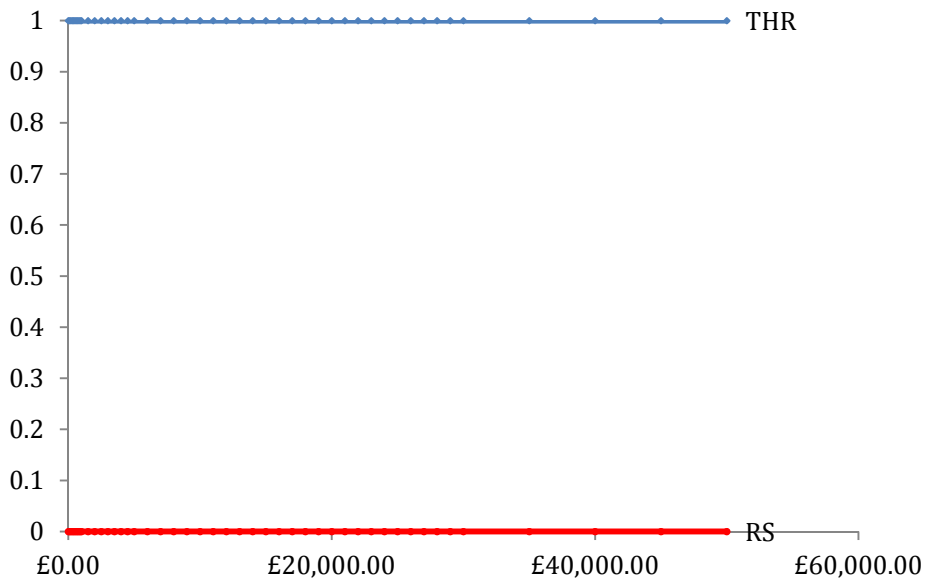
58a: CE plane for 10-year time horizon



58b: CE plane for lifetime horizon



58c: CEAC for 10-year time horizon



58d: CEAC for lifetime horizon

Figure 58. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. RS

10.2.2 The cost-effectiveness analysis results: comparison of THR categories

For 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 91 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 four categories; however, the QALYs were slightly more for Category E than the other four categories. The incremental cost-effectiveness ratio between Category A and 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 scenarios (both deterministic and probabilistic), the mean costs for Category E were slightly lower and the mean QALYs for Category E were slightly higher, than all other four THR categories. Hence, Category E dominated the other four categories.

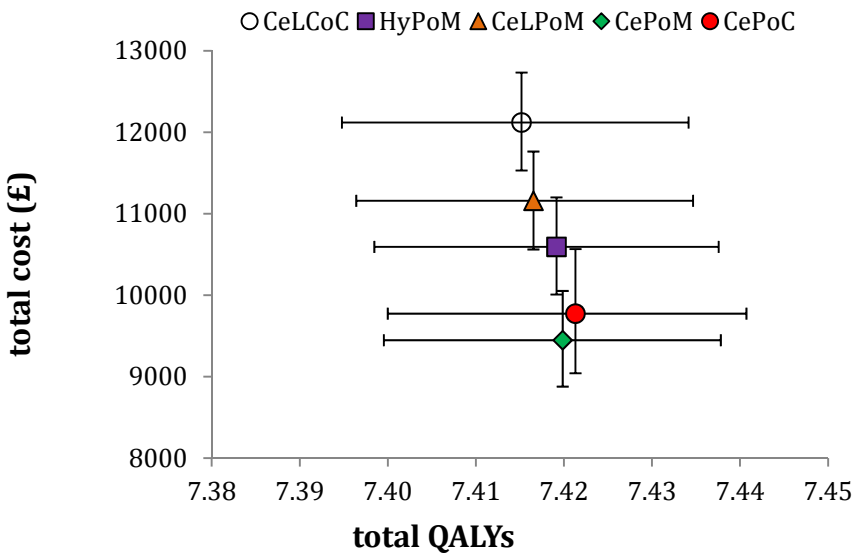
Table 91. Base-case deterministic and probabilistic results for all THR patients using bathtub model

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,444	7.4189	-	-	-	-
E	9,743	7.4207	E v A	299	0.0018	166,217
D	10,588	7.4182	D v E	845	-0.0025	Dominated
B	11,155	7.4156	B v D	567	-0.0026	Dominated
C	12,112	7.4143	C v B	957	-0.0013	Dominated
Deterministic: Lifetime horizon						
E	14,522	14.7909	-	-	-	-
A	14,801	14.7887	A v E	278	-0.0022	Dominated
D	16,040	14.7881	D v A	1,240	-0.0006	Dominated
B	16,804	14.7861	B v D	764	-0.0020	Dominated
C	18,226	14.7845	C v B	1,422	-0.0016	Dominated
Probabilistic: 10-year time horizon						
A	9,449	7.4199	-	-	-	-
E	9,775	7.4213	E v A	326	0.0014	225,225
D	10,594	7.4192	D v E	820	-0.0021	Dominated
B	11,160	7.4165	B v D	566	-0.0026	Dominated
C	12,121	7.4152	C v B	961	-0.0014	Dominated
Probabilistic: Lifetime horizon						
E	14,456	14.7914	-	-	-	-
A	14,740	14.7892	A v E	284	-0.0022	Dominated
D	15,975	14.7885	D v A	1,234	-0.0006	Dominated
B	16,730	14.7866	B v D	755	-0.0019	Dominated
C	18,163	14.7850	C v B	1,432	-0.0016	Dominated

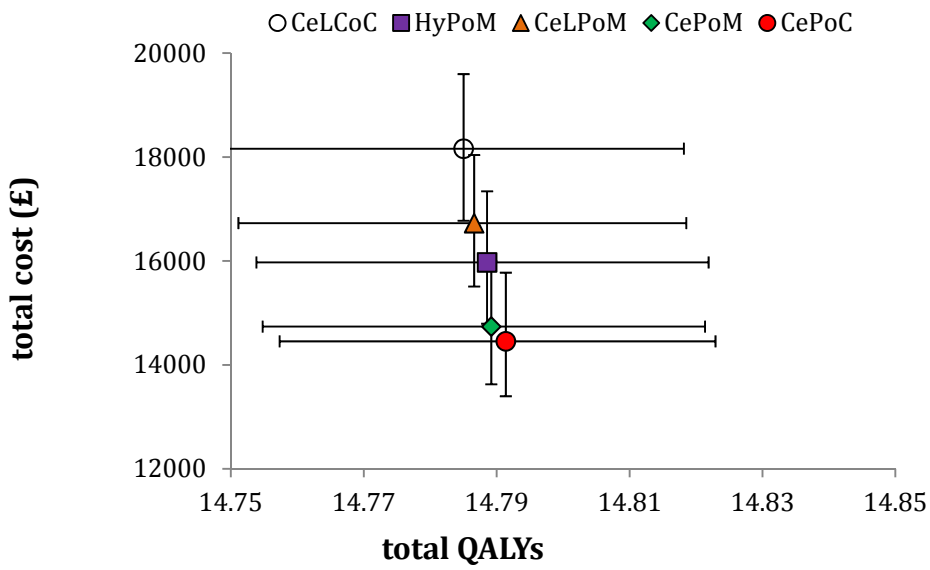
Figure 59 a and b show the cost-effectiveness planes with 95% confidence intervals. For the 10-year time horizon, although Category A is cheaper, Category E generates more QALYs. For the lifetime horizon, Category E is more cost-effective (that is, cheaper and more effective) than the other four categories. Figure 59c and d shows the cost-effectiveness acceptability curves for THR vs. THR using a bathtub model for the two time horizons. For the 10-year time horizon, if the decision maker was willing to pay £20,000 per QALY, Category A was 95% more cost-effective than the other four categories (see Figure 59. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. THR using a bathtub model

c). For the lifetime horizon, if a decision maker is willing to pay anything from £0 to £50,000 per QALY, Category E is over 90% cost-effective. (see Figure 59. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. THR using a bathtub model

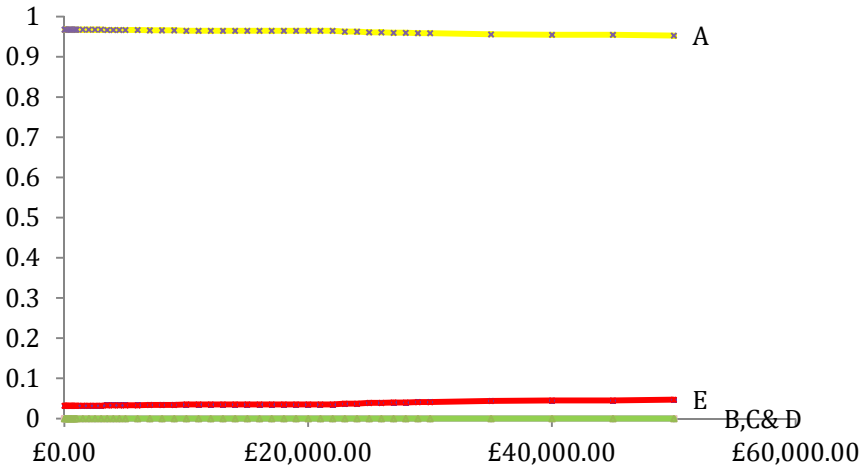
d).



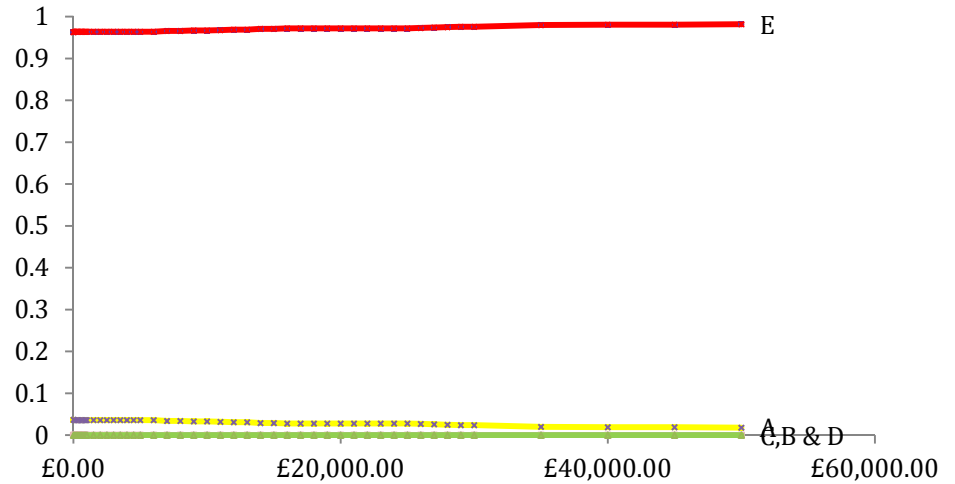
59a: CE plane for 10-year time horizon



59b: CE plane for lifetime horizon



59c: CEAC for 10-year time horizon



59d: CEAC for lifetime horizon

Figure 59. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. THR using a bathtub model

10.3 Sensitivity analysis results

This section presents the results from the deterministic and probabilistic sensitivity analyses.

10.3.1 Subgroup analyses: RS vs. THR

Table 92 and Table 93 shows the deterministic and probabilistic results for RS compared with THR presented separately for males and females 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 compared with 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.

Table 92. Deterministic results for 40, 50 and 60 year old males and female patients

	Age 40		Age 50		Age 60	
	RS	**THR**	**RS**	**THR**	**RS**	**THR**
Women: 10-year time horizon						
Total mean costs £	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	
ICERs (£/QALY)	Dominated		Dominated		Dominated	
Women: Lifetime horizon						
Total mean costs £	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	
ICERs (£/QALY)	Dominated		Dominated		Dominated	
Men: 10-year time horizon						
Total mean costs £	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	
ICERs(£/QALY)	Dominated		Dominated		Dominated	
Men: Lifetime horizon						
Total mean costs £	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 £	9,283		10,672		11,310	
Incremental QALYs	-0.0879		-0.0797		-0.0593	
ICERs (£/QALY)	Dominated		Dominated		Dominated	

Table 93. Probabilistic results for 40, 50 and 60 year old male and female patients

	Age 40		Age 50		Age 60	
	RS	**THR**	**RS**	**THR**	**RS**	**THR**
Women: 10-year time horizon						
Total mean costs £	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	
ICERs (£/QALY)	Dominated		Dominated		Dominated	
Women: Lifetime horizon						
Total mean costs £	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	
ICERs (£/QALY)	Dominated		Dominated		Dominated	
Men: 10-year time horizon						
Total mean costs £	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	
ICERs (£/QALY)	Dominated		Dominated		Dominated	
Men: Lifetime horizon						
Total mean costs £	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 £	9,231		10,635		11,292	
Incremental QALYs	-0.0880		-0.0799		-0.0595	
ICERs (£/QALY)	Dominated		Dominated		Dominated	

The results from Table 92 and Table 93 are reflected in the cost-effectiveness planes and the cost-effectiveness acceptability curves (Figure 60 and Figure 61).

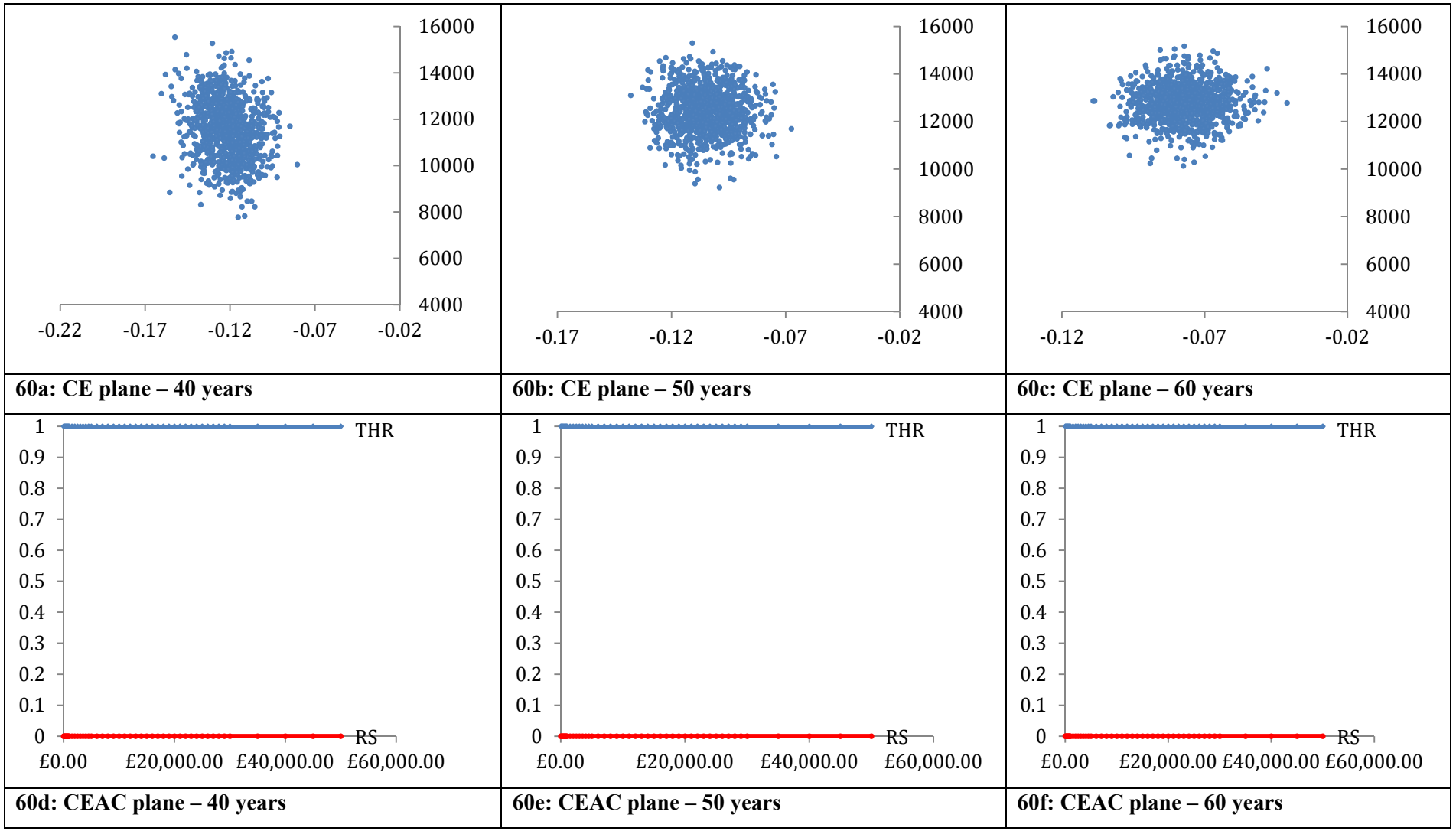
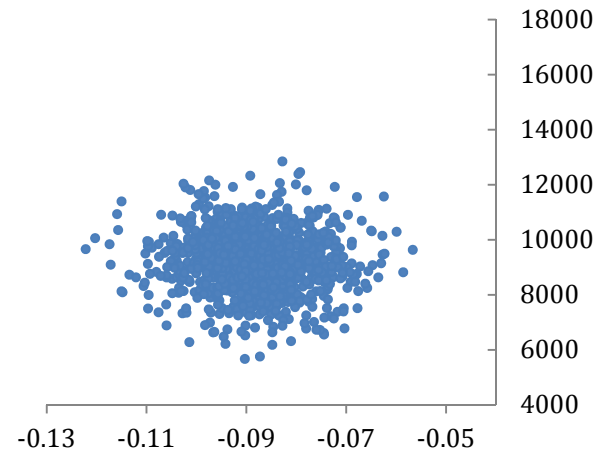
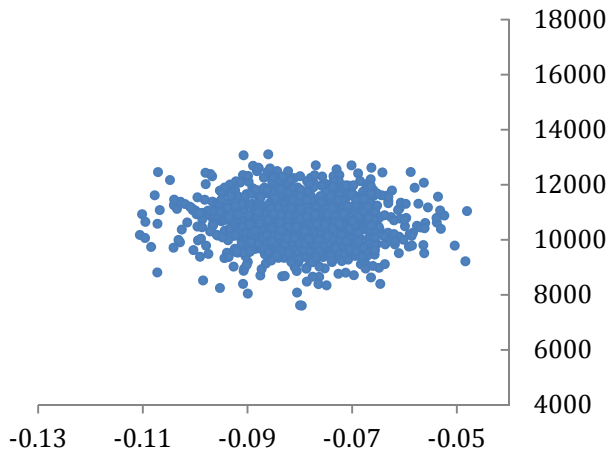


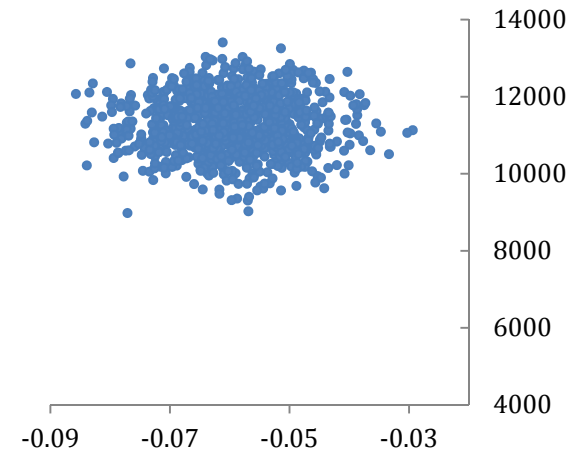
Figure 60. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. RS for females by age group (lifetime horizon)



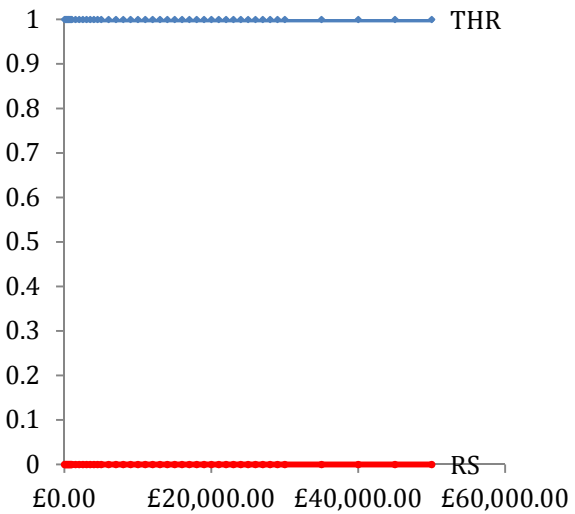
61a: CE plane – 40 years



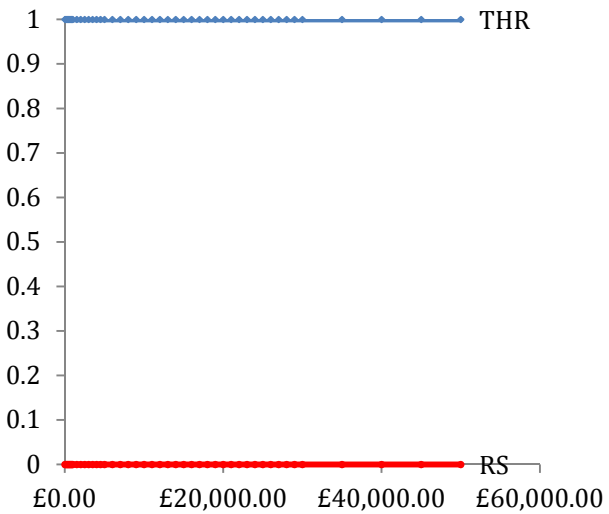
61b: CE plane – 50 years



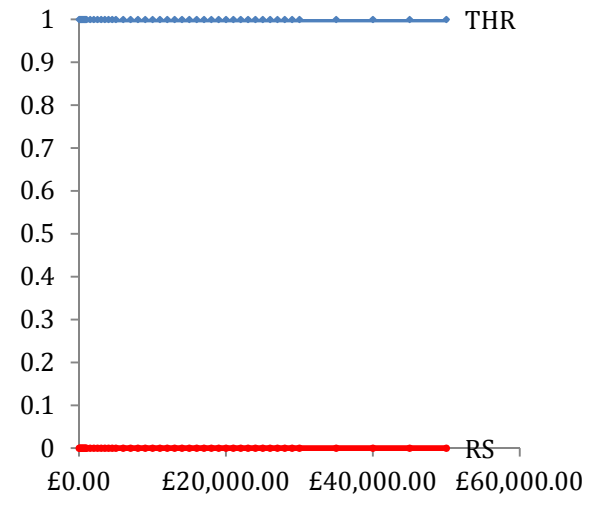
61c: CE plane – 60 years



61d: CEAC plane – 40 years



61e: CEAC plane – 50 years



61f: CEAC plane – 60 years

Figure 61. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. RS for males by age group (lifetime horizon)

10.3.2 Subgroup analyses: THR vs. THR (over 65's)

For the different THR categories split by age and gender, deterministic and probabilistic results over the 10-year time horizon are shown below (

Table 94 and Table 95) along with the corresponding ICERs (where appropriate). For both men and women in the 70 years and 80 years age groups, although Category A was cheaper, Category E was more effective.

Table 94. Deterministic results for males and females over 65 years of age for a 10-year time horizon

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age group 70						
Women: 10-year time horizon						
A	9,047	6.8159	-	-	-	-
E	9,364	6.8173	E v A	317	0.0014	231,970
D	10,134	6.8160	D v E	770	-0.0013	Dominated
B	10,586	6.8150	B v D	452	-0.0010	Dominated
C	11,427	6.8151	C v B	841	0.0001	5,773,991
A	9,047	6.8159	-	-	-	-
E	9,364	6.8173	E v A	317	0.0014	231,970
C	11,427	6.8151	C v E	2,063	-0.0022	Dominated
Men: 10-year time horizon						
A	8,900	6.8903	-	-	-	-
E	9,238	6.8915	E v A	338	0.0012	281,096
D	10,028	6.8898	D v E	790	-0.0016	Dominated
B	10,506	6.8885	B v D	478	-0.0013	Dominated
C	11,451	6.8874	C v B	944	-0.0011	Dominated
Age 80						
Women: 10-year time horizon						
A	8,175	5.1980	-	-	-	-
E	8,495	5.1984	E v A	320	0.0004	803,012
D	9,263	5.1981	D v E	768	-0.0003	Dominated
B	9,829	5.1975	B v D	566	-0.0006	Dominated
C	10,681	5.1975	C v B	851	-0.0000	Dominated
Men: 10-year time horizon						
A	8,035	5.0689	-	-	-	-
E	8,464	5.0690	E v A	429	0.0000	12,763,540
D	9,138	5.0689	D v E	673	-0.0001	Dominated
B	9,752	5.0679	B v D	615	-0.0010	Dominated
C	10,695	5.0675	C v B	942	-0.0004	Dominated

Table 95. Probabilistic results for males and females over 65 years of age for a 10-year time horizon

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age group 70						
Women: 10-year time horizon						
A	9,046	6.8161	-	-	-	-
E	9,362	6.8174	E v A	316	0.0014	229,667
D	10,139	6.8160	D v E	777	-0.0014	Dominated
B	10,591	6.8151	B v D	452	-0.0010	Dominated
C	11,425	6.8153	C v B	834	0.0002	3,786,953
A	9,046	6.8161	-	-	-	-
E	9,362	6.8174	E v A	316	0.0014	229,667
C	11,425	6.8153	C v E	2,063	-0.0022	Dominated
Men: 10-year time horizon						
A	8,891	6.8905	-	-	-	-
E	9,268	6.8912	E v A	377	0.0007	512,560
D	10,027	6.8900	D v E	759	-0.0013	Dominated
B	10,503	6.8886	B v D	476	-0.0013	Dominated
C	11,508	6.8868	C v B	1,005	-0.0018	Dominated
Age 80						
Women: 10-year time horizon						
A	8,170	5.1985	-	-	-	-
E	8,490	5.1989	E v A	320	0.0004	804,850
D	9,260	5.1985	D v E	770	-0.0003	Dominated
B	9,828	5.1979	B v D	568	-0.0006	Dominated
C	10,675	5.1979	C v B	846	0.0000	1,573,299,053
A	8,170	5.1985	-	-	-	-
E	8,490	5.1989	E v A	320	0.0004	804,850
C	10,675	5.1979	C v E	2,184	-0.0009	Dominated
Men: 10-year time horizon						
A	8,029	5.0687	-	-	-	-
E	8,501	5.0686	E v A	472	-0.0002	Dominated
D	9,140	5.0687	D v E	639	0.0001	8,491,620
B	9,753	5.0676	B v D	614	-0.0010	Dominated
C	10,768	5.0669	C v B	1,015	-0.0007	Dominated
A	8,029	5.0687	-	-	-	-
D	9,140	5.0687	D v A	1,110	-0.0001	Dominated

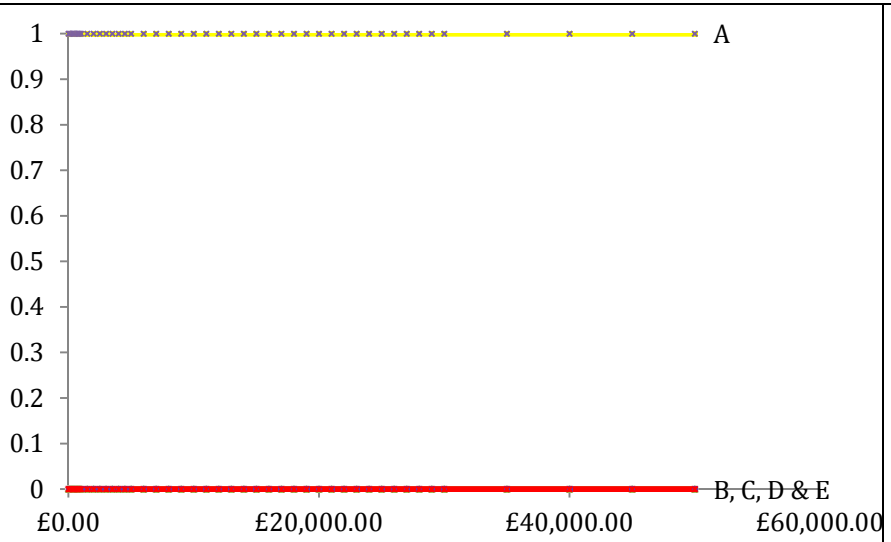
Table 96 and Table 97 show the deterministic and probabilistic results for males and females over the age of 65 (70 and 80 years) for the lifetime horizon along with the corresponding ICERs (where appropriate). For both men and women in the 70 years age group, although Category A was cheaper, Category E was more effective; for women over the age of 80, Category A was cheaper and Category D generated more QALYs, and for men over the age of 80, Category A was cheaper, although 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 62.

Table 96. Deterministic results for males and females over 65 years of age for a lifetime horizon

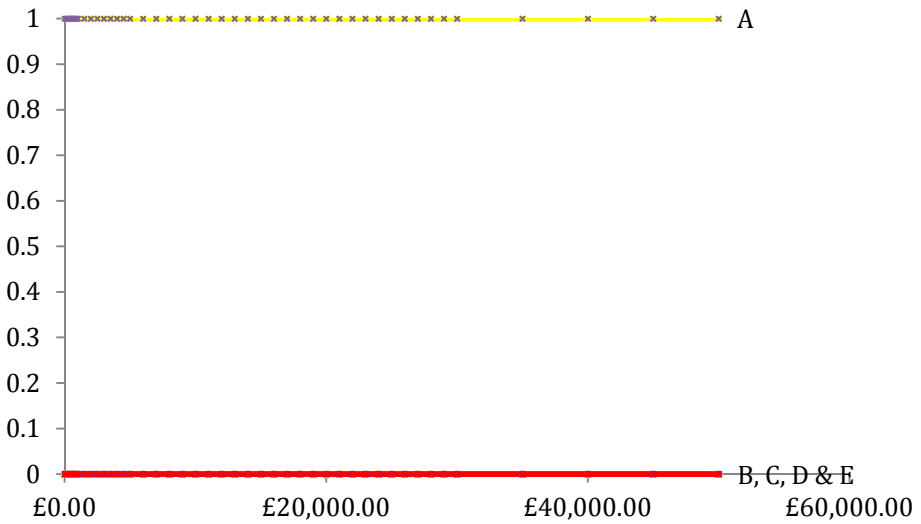
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age group 70						
Women: Lifetime horizon						
A	10,635	9.4317	-	-	-	-
E	10,916	9.4318	E v A	281	0.0001	3,208,305
D	11,694	9.4316	D v E	778	-0.0001	Dominated
B	12,160	9.4315	B v D	466	-0.0001	Dominated
C	13,005	9.4316	C v B	845	0.0000	23,645,296
A	10,635	9.4317	-	-	-	-
E	10,916	9.4318	E v A	281	0.0001	3,208,305
C	13,005	9.4316	C v E	2,090	-0.0002	Dominated
Men: Lifetime horizon						
A	10,111	8.9914	-	-	-	-
E	10,428	8.9916	E v A	317	0.0002	1,424,339
D	11,247	8.9913	D v E	819	-0.0003	Dominated
B	11,738	8.9911	B v D	492	-0.0003	Dominated
C	12,712	8.9909	C v B	973	-0.0002	Dominated
Age 80						
Women: Lifetime horizon						
A	8,688	6.0572	-	-	-	-
E	8,993	6.0573	E v A	305	0.0002	1,911,863
D	9,768	6.0574	D v E	774	0.0000	15,988,179
B	10,350	6.0573	B v D	583	0.0000	Dominated
C	11,204	6.0573	C v B	854	-0.0001	Dominated
Men: Lifetime horizon						
A	8,391	5.6873	-	-	-	-
E	8,820	5.6873	E v A	429	0.0000	118,964,663
D	9,494	5.6873	D v E	674	0.0000	Dominated
B	10,123	5.6868	B v D	629	-0.0005	Dominated
C	11,075	5.6866	C v B	952	-0.0003	Dominated

Table 97. Probabilistic results for males and females over 65 years of age for a lifetime horizon

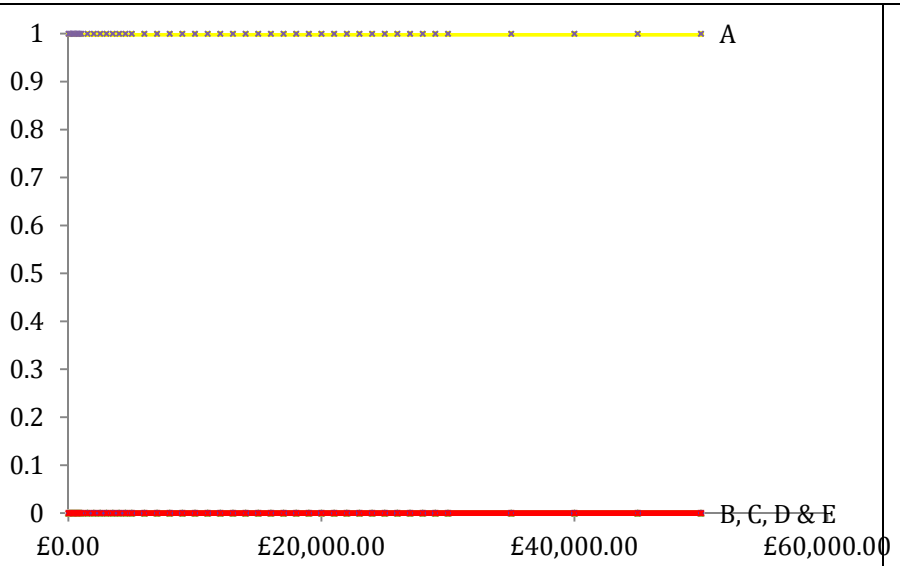
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age group 70						
Women: Lifetime horizon						
A	10,636	9.4314	-	-	-	-
E	10,919	9.4315	E v A	282	0.0001	3,168,484
D	11,708	9.4314	D v E	789	-0.0001	Dominated
B	12,168	9.4313	B v D	460	-0.0001	Dominated
C	13,006	9.4313	C v B	838	0.0000	20,570,154
A	10,636	9.4314	-	-	-	-
E	10,919	9.4315	E v A	282	0.0001	3,168,484
C	13,006	9.4313	C v E	2,088	-0.0002	Dominated
Men: Lifetime horizon						
A	10,099	8.9914	-	-	-	-
E	10,458	8.9915	E v A	359	0.0002	2,342,245
D	11,243	8.9913	D v E	786	-0.0002	Dominated
B	11,732	8.9910	B v D	489	-0.0003	Dominated
C	12,778	8.9907	C v B	1,046	-0.0003	Dominated
Age 80						
Women: Lifetime horizon						
A	8,690	6.0579	-	-	-	-
E	8,995	6.0581	E v A	305	0.0002	1,964,904
D	9,774	6.0582	D v E	779	0.0001	15,297,263
B	10,356	6.0581	B v D	582	0.0000	Dominated
C	11,205	6.0580	C v B	850	-0.0001	Dominated
Men: Lifetime horizon						
A	8,395	5.6873	-	-	-	-
E	8,866	5.6872	E v A	471	-0.0001	Dominated
D	9,508	5.6873	D v E	643	0.0001	12,759,024
B	10,133	5.6868	B v D	625	-0.0004	Dominated
C	11,164	5.6864	C v B	1,031	-0.0004	Dominated
A	8,395	5.6873	-	-	-	-
D	9,508	5.6873	D v A	1,114	-0.0001	Dominated



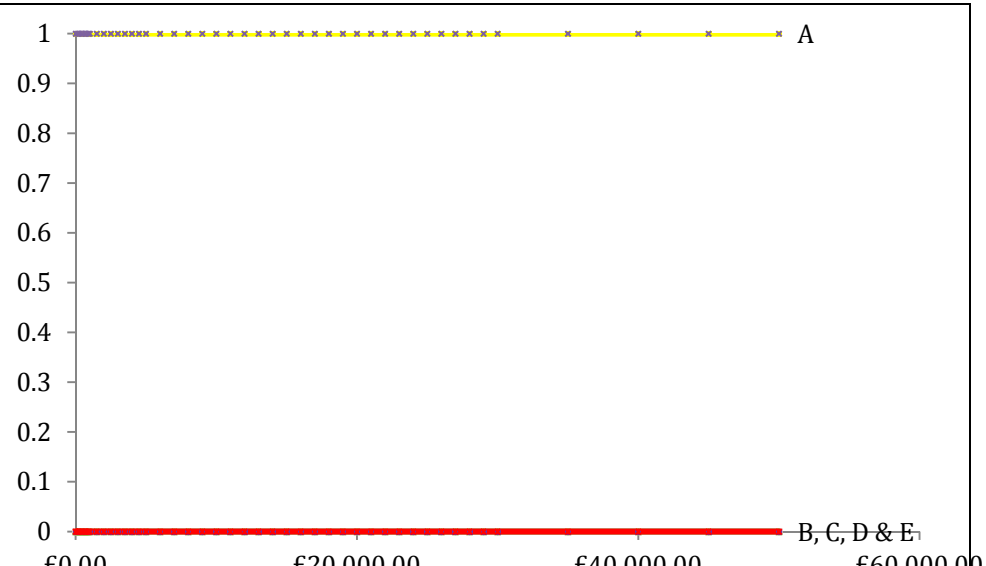
62a: CEAC for lifetime horizon over 70s females



62b: CEAC for lifetime horizon over 80s females



62c: CEAC for lifetime horizon over 70s males



62d: CEAC for lifetime horizon over 80s males

Figure 62. Cost-effectiveness acceptability curves for THR vs. THR (over 65s)

10.3.3 Subgroup analyses: THR vs. THR (under 65's)

For the different THR categories split by age and gender, deterministic and probabilistic results over the 10-year time horizon are shown below (Table 98 and Table 99) along with the corresponding ICERs (where appropriate). For men in the age group 40, 50 and 60 years age groups, although Category A was cheaper, Category E was more effective.

Table 98. Deterministic results for males under 65 years of age for 10-year time horizon

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Men: 10-year time horizon						
A	10,097	7.3299	-	-	-	-
E	10,289	7.3330	E v A	192	0.0031	62,892
D	11,398	7.3274	D v E	1,109	-0.0056	Dominated
B	11,742	7.3277	B v D	344	0.0004	947,877
C	12,452	7.3294	C v B	711	0.0016	434,139
A	10,097	7.3299	-	-	-	-
E	10,289	7.3330	E v A	192	0.0031	62,892
B	11,742	7.3277	B v E	1,452	-0.0052	Dominated
C	12,452	7.3294	C v B	710	0.0016	434,139
A	10,097	7.3299	-	-	-	-
E	10,289	7.3330	E v A	192	0.0031	62,892
C	12,452	7.3294	C v E	2,163	-0.0036	Dominated
Age 50						
Men: 10-year time horizon						
A	9,833	7.5230	-	-	-	-
E	9,991	7.5270	E v A	157	0.0039	40,250
D	11,133	7.5202	D v E	1,143	-0.0068	Dominated
B	11,647	7.5182	B v D	514	-0.0020	Dominated
C	12,274	7.5213	C v B	627	0.0030	205,546
A	9,833	7.5230	-	-	-	-
E	9,991	7.5270	E v A	157	0.0039	40,250
C	12,274	7.5213	C v E	2,283	-0.0057	Dominated
Age 60						
Men: 10-year time horizon						
A	9,529	7.5085	-	-	-	-
E	9,685	7.5126	E v A	156	0.0042	37,466
D	10,819	7.5056	D v E	1,134	-0.0071	Dominated
B	11,460	7.5016	B v D	642	-0.0040	Dominated
C	12,025	7.5057	C v B	565	0.0042	135,491
A	9,529	7.5085	-	-	-	-
E	9,685	7.5126	E v A	156	0.0042	37,466
C	12,025	7.5057	C v E	2,340	-0.0069	Dominated

Table 99. Probabilistic results for males under 65 years of age for 10-year time horizon

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Men: 10-year time horizon						
A	10,178	7.3290	-	-	-	-
E	10,390	7.3318	E v A	212	0.0028	74,551
D	11,623	7.3247	D v E	1,233	-0.0071	Dominated
B	11,837	7.3266	B v D	214	0.0019	112,217
C	12,474	7.3292	C v B	637	0.0025	253,807
A	10,178	7.3290	-	-	-	-
E	10,390	7.3318	E v A	212	0.0028	74,551
B	11,837	7.3266	B v E	1,447	-0.0052	Dominated
C	12,474	7.3292	C v B	637	0.0025	253,807
A	10,178	7.3290	-	-	-	-
E	10,390	7.3318	E v A	212	0.0028	74,551
C	12,474	7.3292	C v E	2,084	-0.0027	Dominated
Age 50						
Men: 10-year time horizon						
A	9,835	7.5227	-	-	-	-
E	10,021	7.5262	E v A	187	0.0035	52,927
D	11,172	7.5193	D v E	1,151	-0.0069	Dominated
B	11,662	7.5177	B v D	490	-0.0016	Dominated
C	12,284	7.5208	C v B	622	0.0031	199,704
A	9,835	7.5227	-	-	-	-
E	10,021	7.5262	E v A	187	0.0035	52,927
C	12,284	7.5208	C v E	2,263	-0.0054	Dominated
Age 60						
Men: 10-year time horizon						
A	9,529	7.5091	-	-	-	-
E	9,685	7.5132	E v A	157	0.0041	37,843
D	10,815	7.5062	D v E	1,130	-0.0070	Dominated
B	11,465	7.5021	B v D	650	-0.0041	Dominated
C	12,028	7.5063	C v B	564	0.0042	134,913
A	9,529	7.5091	-	-	-	-
E	9,685	7.5132	E v A	157	0.0041	37,843
C	12,028	7.5063	C v E	2,343	-0.0069	Dominated

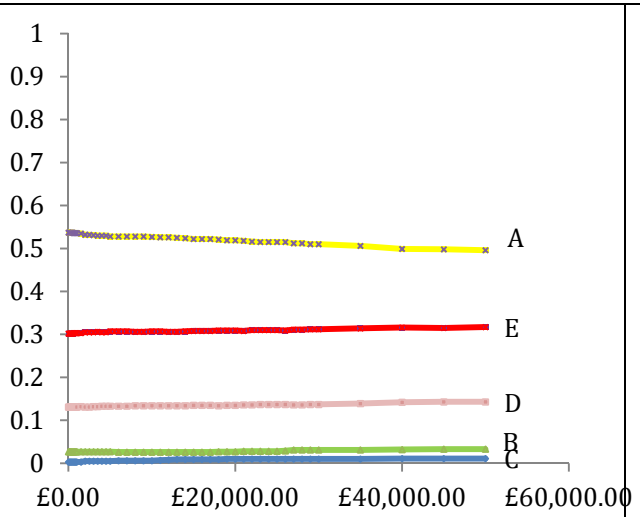
For men deterministic and probabilistic results over the lifetime horizon are shown below (Table 100 and Table 101) along with the corresponding ICERs (where appropriate). In the age group 40 years, Category A dominated the other four categories; and for men in the 50 years and 60 years age groups, Category A was cheaper, however, Category E was more effective. Figure 63 show the corresponding CEACs**.**

Table 100. Deterministic results for males under 65 years of age for a lifetime horizon

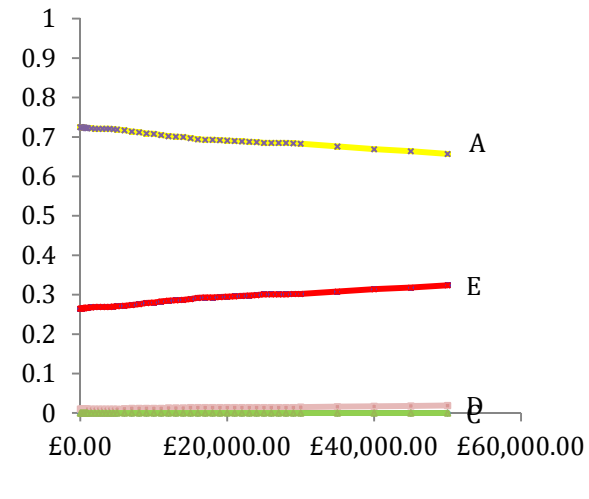
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Men: Lifetime horizon						
A	18,350	16.6684	-	-	-	-
E	19,351	16.6677	E v A	1,000	-0.0008	Dominated
D	20,572	16.6625	D v E	1,222	-0.0052	Dominated
C	21,270	16.6656	C v D	697	0.0032	219,152
B	21,712	16.6593	B v C	442	-0.0063	Dominated
A	18,350	16.6684	-	-	-	-
C	21,270	16.6656	C v A	2,919	-0.0028	Dominated
Age 50						
Men: Lifetime horizon						
A	15,579	14.8116	-	-	-	-
E	15,998	14.8132	E v A	419	0.0016	257,281
D	17,560	14.8081	D v E	1,561	-0.0052	Dominated
C	18,579	14.8090	C v D	1,020	0.0010	1,059,918
B	19,016	14.8047	B v C	437	-0.0044	Dominated
A	15,579	14.8116	-	-	-	-
E	15,998	14.8132	E v A	419	0.0016	257,281
C	18,579	14.8090	C v E	2,581	-0.0042	Dominated
Age 60						
Men: Lifetime horizon						
A	12,929	12.2177	-	-	-	-
E	13,082	12.2192	E v A	153	0.0014	105,773
D	14,606	12.2158	D v E	1,524	-0.0034	Dominated
C	15,819	12.2158	C v D	1,213	0.0001	14,646,830
B	16,011	12.2132	B v C	192	-0.0026	Dominated
A	12,929	12.2177	-	-	-	-
E	13,082	12.2192	E v A	153	0.0014	105,773
C	15,819	12.2158	C v E	2,737	-0.0033	Dominated

Table 101. Probabilistic results for males under 65 years of age for a lifetime horizon

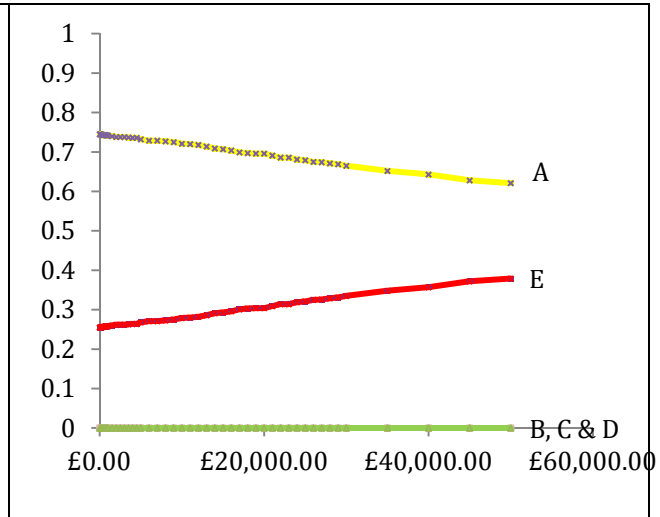
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Men: Lifetime horizon						
A	18,556	16.6662	-	-	-	-
E	19,587	16.6651	E v A	1,031	-0.0011	Dominated
D	21,069	16.6577	D v E	1,481	-0.0073	Dominated
C	21,304	16.6646	C v D	235	0.0068	34,383
B	21,877	16.6570	B v C	573	-0.0076	Dominated
A	18,556	16.6662	-	-	-	-
C	21,304	16.6646	C v A	2,748	-0.0016	Dominated
Age 50						
Men: Lifetime horizon						
A	15,626	14.8108	-	-	-	-
E	16,071	14.8124	E v A	444	0.0016	279,122
D	17,608	14.8074	D v E	1,538	-0.0051	Dominated
C	18,581	14.8085	C v D	973	0.0012	843,588
B	19,032	14.8041	B v C	451	-0.0044	Dominated
A	15,626	14.8108	-	-	-	-
E	16,071	14.8124	E v A	444	0.0016	279,122
C	18,581	14.8085	C v E	2,511	-0.0039	Dominated
Age 60						
Men: Lifetime horizon						
A	12,957	12.2174	-	-	-	-
E	13,113	12.2188	E v A	156	0.0014	109,045
D	14,617	12.2155	D v E	1,503	-0.0033	Dominated
C	15,831	12.2155	C v D	1,215	0.0001	15,339,725
B	16,029	12.2128	B v C	198	-0.0027	Dominated
A	12,957	12.2174	-	-	-	-
E	13,113	12.2188	E v A	156	0.0014	109,045
C	15,831	12.2155	C v E	2,718	-0.0033	Dominated



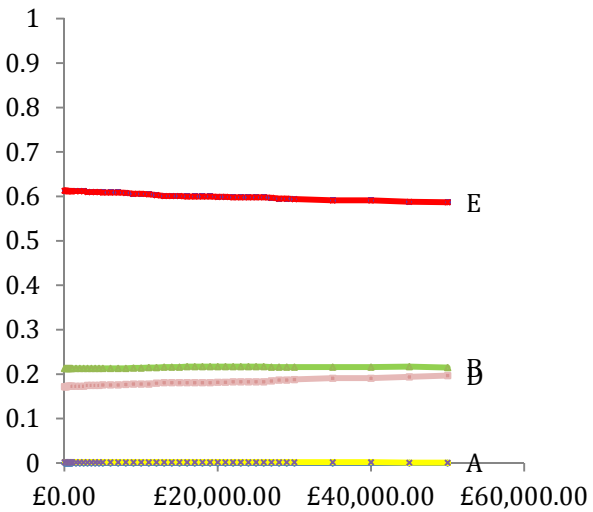
63a: CEAC – males 40 years



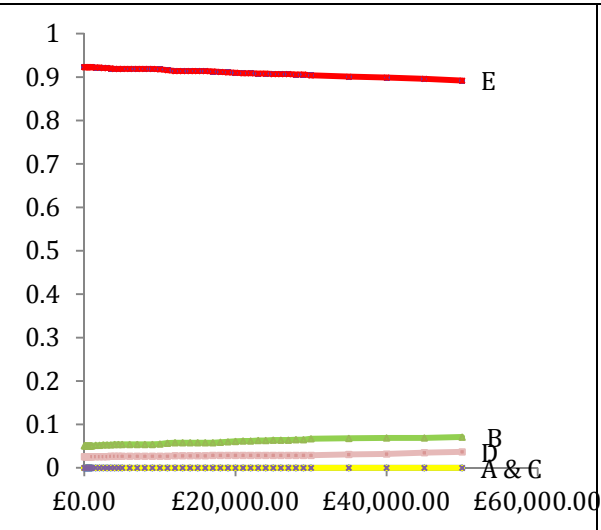
63b: CEAC – males 50 years



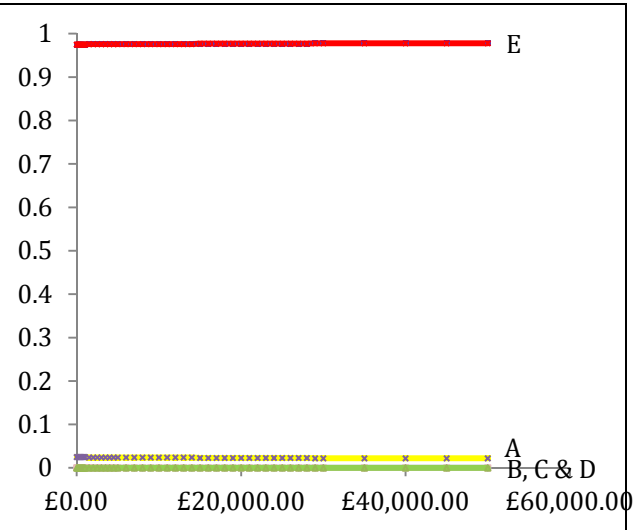
63c: CEAC – males 60 years



63d: CEAC – females 40 years



63e: CEAC – females 50 years



63f: CEAC – females 60 years

Figure 63. Cost-effectiveness acceptability curves for THR vs. THR for females and males by age group (lifetime horizon)

For women, deterministic and probabilistic results over the 10-year time horizon are shown below (Table 102 and Table 103) along with the corresponding ICERs (where appropriate). In the age groups 40 and 50 years, Category E dominated the other four categories; for women in the age group 60 years although Category A was cheaper, Category E was more effective.

Table 102. Deterministic results for females under 65 years of age for a 10-year time horizon

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Women: 10-year time horizon						
E	10,064	7.1950	-	-	-	-
A	10,437	7.1856	A v E	374	-0.0094	Dominated
D	10,805	7.1940	D v A	368	0.0084	43,732
B	11,540	7.1897	B v D	735	-0.0043	Dominated
C	12,381	7.1898	C v B	841	0.0001	7,772,228
E	10,064	7.1950	-	-	-	-
D	10,805	7.1940	D v E	742	-0.0010	Dominated
C	12,381	7.1898	C v D	1,575	-0.0042	Dominated
Age 50						
Women: 10-year time horizon						
E	9,978	7.3423	-	-	-	-
A	10,035	7.3359	A v E	57	-0.0064	Dominated
D	10,802	7.3401	D v A	766	0.0042	181,105
B	11,355	7.3376	B v D	553	-0.0025	Dominated
C	12,251	7.3371	C v B	896	-0.0006	Dominated
E	9,978	7.3423	-	-	-	-
D	10,802	7.3401	D v E	823	-0.0022	Dominated
Age 60						
Women: 10-year time horizon						
A	9,670	7.4075	-	-	-	-
E	9,846	7.4112	E v A	176	0.0037	48,110
D	10,743	7.4078	D v E	897	-0.0033	Dominated
B	11,137	7.4073	B v D	394	-0.0005	Dominated
C	12,074	7.4061	C v B	937	-0.0012	Dominated

Table 103. Probabilistic results for females under 65 years of age for a 10-year time horizon

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Women: 10-year time horizon						
E	9,983	7.1954	-	-	-	-
A	10,502	7.1843	A v E	520	-0.0011	Dominated
D	10,967	7.1916	D v A	464	0.0073	63,507
B	11,630	7.1881	B v D	663	-0.0035	Dominated
C	12,405	7.1890	C v B	775	0.0009	889,457
E	9,983	7.1954	-	-	-	-
D	10,967	7.1916	D v E	984	-0.0038	Dominated
C	12,405	7.1890	C v D	1,438	-0.0027	Dominated
Age 50						
Women: 10-year time horizon						
E	9,936	7.3426	-	-	-	-
A	10,049	7.3355	A v E	113	-0.0071	Dominated
D	10,849	7.3393	D v A	800	0.0038	209,865
B	11,384	7.3371	B v D	535	-0.0022	Dominated
C	12,253	7.3368	C v B	869	-0.0002	Dominated
E	9,936	7.3426	-	-	-	-
D	10,849	7.3393	D v E	913	-0.0033	Dominated
Age 60						
Women: 10-year time horizon						
A	9,673	7.4075	-	-	-	-
E	9,849	7.4111	E v A	176	0.0037	48,113
D	10,749	7.4077	D v E	900	-0.0034	Dominated
B	11,147	7.4072	B v D	398	-0.0006	Dominated
C	12,075	7.4061	C v B	928	-0.0010	Dominated

For women deterministic and probabilistic results over the lifetime horizon are shown below (Table 104 and Table 105) along with the corresponding ICERs (where appropriate). In the age groups 40, 50 and 60 years, Category E dominated the other four categories (i.e., Category E was cheaper and more effective). Figure 63 earlier shows the corresponding CEACs**.**

Table 104. Deterministic results for females under 65 years of age for a lifetime horizon

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Women: Lifetime horizon						
E	18,647	16.8374	-	-	-	-
B	18,814	16.8361	B v E	167	-0.0013	Dominated
D	20,033	16.8340	D v B	1,218	-0.0020	Dominated
A	21,595	16.8180	A v D	1,562	-0.0160	Dominated
C	21,886	16.8289	C v A	291	0.0109	26,657
E	18,647	16.8374	-	-	-	-
C	21,886	16.8289	C v E	3,238	-0.0085	Dominated
Age 50						
Women: Lifetime horizon						
E	16,426	15.1069	-	-	-	-
B	16,923	15.1053	B v E	497	-0.0016	Dominated
A	17,854	15.1003	A v B	931	-0.0050	Dominated
D	18,024	15.1042	D v A	170	0.0039	43,755
C	19,366	15.1022	C v D	1,342	-0.0020	Dominated
E	16,426	15.1069	-	-	-	-
D	18,024	15.1042	D v E	1,598	-0.0027	Dominated
Age 60						
Women: Lifetime horizon						
E	14,026	12.6801	-	-	-	-
A	14,343	12.6785	A v E	317	-0.0016	Dominated
B	14,844	12.6798	B v A	501	0.0013	398,183
D	15,599	12.6787	D v B	755	-0.0011	Dominated
C	16,655	12.6779	C v D	1,056	-0.0008	Dominated
E	14,026	12.6801	-	-	-	-
B	14,844	12.6798	B v E	818	-0.0004	Dominated

Table 105. Probabilistic results for females under 65 years of age for a lifetime horizon

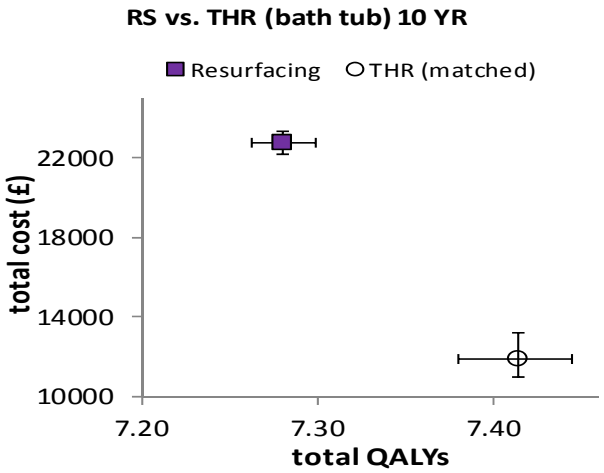
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Age 40						
Women: Lifetime horizon						
E	18,179	16.8404	-	-	-	-
B	19,050	16.8351	B v E	871	-0.0053	Dominated
D	20,368	16.8317	D v B	1,318	-0.0034	Dominated
A	21,704	16.8178	A v D	1,335	-0.0139	Dominated
C	21,959	16.8291	C v A	255	0.0113	22,538
E	18,179	16.8404	-	-	-	-
C	21,959	16.8291	C v E	3,780	-0.0113	Dominated
Age 50						
Women: Lifetime horizon						
E	16,425	15.1072	-	-	-	-
B	16,980	15.1048	B v E	735	-0.0024	Dominated
A	17,875	15.0999	A v B	895	-0.0049	Dominated
D	18,135	15.1035	D v A	259	0.0036	71,800
C	19,379	15.1018	C v D	1,245	-0.0017	Dominated
E	16,425	15.1072	-	-	-	-
D	18,135	15.1035	D v E	1,889	-0.0037	Dominated
Age 60						
Women: Lifetime horizon						
E	14,031	12.6798	-	-	-	-
A	14,359	12.6781	A v E	328	-0.0017	Dominated
B	14,873	12.6793	B v A	514	0.0012	414,092
D	15,624	12.6782	D v B	751	-0.0011	Dominated
C	16,673	12.6774	C v D	1,048	-0.0008	Dominated
E	14,031	12.6798	-	-	-	-
B	14,873	12.6793	B v E	842	-0.0004	Dominated

10.3.4 Sensitivity analyses: time to revision (bathtub model adjusted for age and gender)

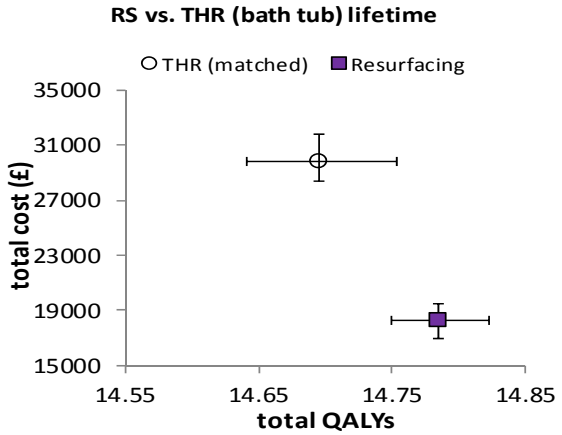
Table 106 below shows the deterministic and probabilistic results for all patients using a bathtub model adjusted for age and gender. Following in line with the base-case analysis, RS was dominated by THR for all time horizons (that is, THR was cheaper and more effective, than RS). The corresponding cost-effectiveness planes and CEACs are shown in Figure 64.

Table 106. Deterministic and probabilistic results for all patients using bathtub model adjusted for age and gender

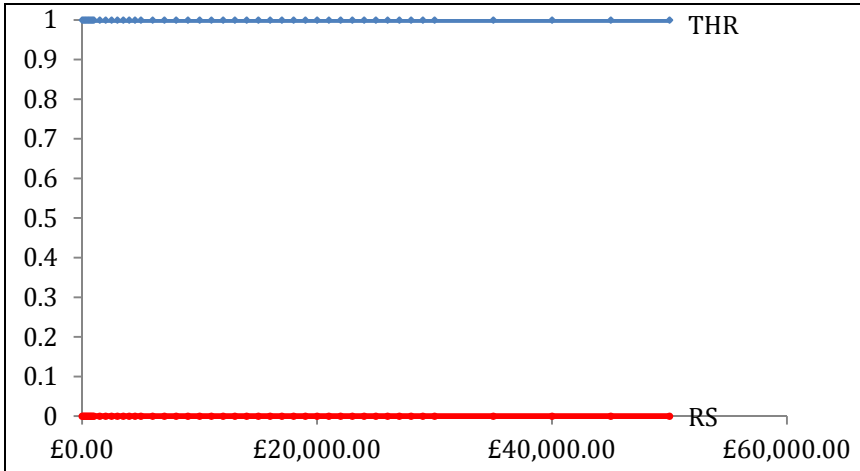
	RS	THR
Deterministic - 10-year time horizon		
Total mean costs £	22,560	11,899
Total mean QALYs	7.2824	7.4144
Incremental cost £	10,661	
Incremental QALYs	-0.1320	
ICERs (£/QALY)	Dominated	
Deterministic - Lifetime horizon		
Total mean costs £	29,664	18,254
Total mean QALYs	14.6964	14.7843
Incremental cost £	11,410	
Incremental QALYs	-0.0879	
ICERs (£/QALY)	Dominated	
Probabilistic - 10-year time horizon		
Total mean costs £	22,729	11,912
Total mean QALYs	7.2804	7.4141
Incremental cost £	10,817	
Incremental QALYs	-0.1337	
ICERs (£/QALY)	Dominated	
Probabilistic - Lifetime horizon		
Total mean costs £	29,836	18,268
Total mean QALYs	14.6958	14.7845
Incremental cost £	11,568	
Incremental QALYs	-0.0887	
ICERs (£/QALY)	Dominated	



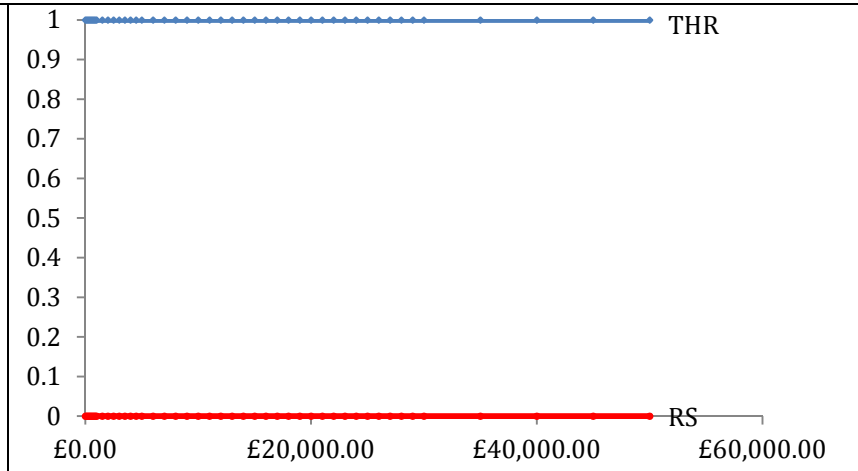
64a. CE Plane for 10-year time horizon



64c. CE Plane for lifetime horizon



64b. CEAC for 10-year time horizon



64d. CEAC for lifetime horizon

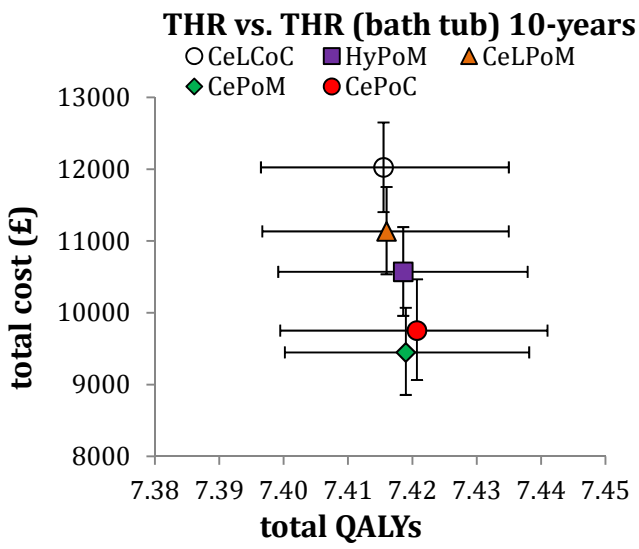
Figure 64. Cost-effectiveness acceptability curves for RS vs. THR age and gender adjusted using a bathtub model

Table 107 shows the deterministic and probabilistic results for all THR patients using a bathtub model adjusted for age and gender. As in the base-case analysis, for the 10-year time horizon (both deterministic and probabilistic) Category A was cheaper than all four categories; however Category E conferred slightly more QALYs than the other four categories. The incremental cost-effectiveness ratio between Category A and Category E was £127,420 per QALY gained for the deterministic analysis and £176,776 per QALY gained for the probabilistic analysis.

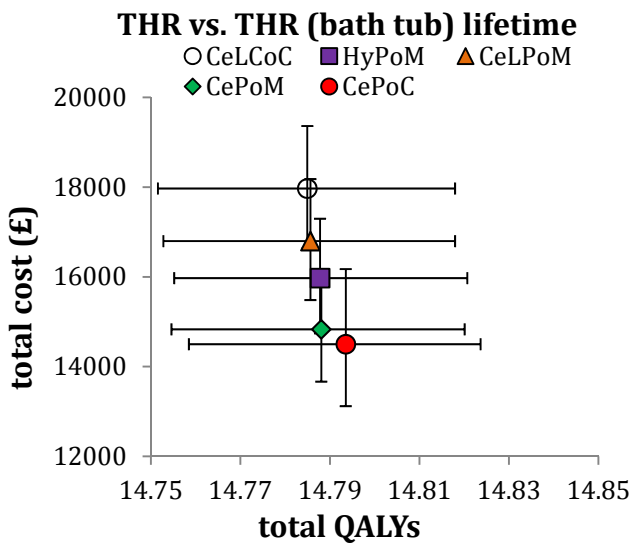
When looking at the lifetime scenarios (both deterministic and probabilistic), mean costs for Category E were slightly lower and the mean QALYs for Category E were slightly higher, than all other four THR categories. Hence, Category E dominated the other four categories. The corresponding cost-effectiveness planes and CEACs are shown in Figure 65.

Table 107. Deterministic and probabilistic results for all THR patients using bathtub model adjusted for age and gender

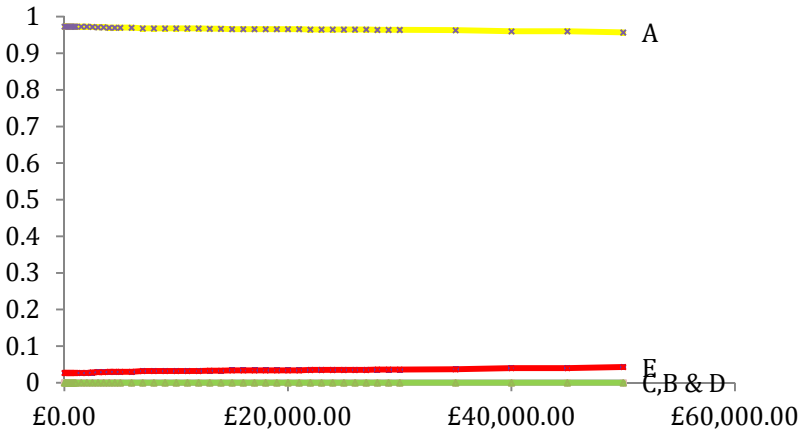
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,458	7.4187	-	-	-	-
E	9,731	7.4208	E v A	273	0.0021	127,420
D	10,578	7.4183	D v E	846	-0.0025	Dominated
B	11,147	7.4157	B v D	569	-0.0027	Dominated
C	12,035	7.4152	C v B	888	-0.0004	Dominated
Deterministic: Lifetime horizon						
E	14,533	14.7909	-	-	-	-
A	14,817	14.7886	A v E	283	-0.0023	Dominated
D	15,965	14.7883	D v A	1,148	-0.0003	Dominated
B	16,784	14.7862	B v D	819	-0.0021	Dominated
C	17,963	14.7854	C v B	1,180	-0.0007	Dominated
Probabilistic: 10-year time horizon						
A	9,449	7.4190	-	-	-	-
E	9,754	7.4207	E v A	304	0.0017	176,776
D	10,572	7.4186	D v E	818	-0.0021	Dominated
B	11,135	7.4160	B v D	563	-0.0026	Dominated
C	12,027	7.4155	C v B	891	-0.0005	Dominated
Probabilistic: Lifetime horizon						
E	13,954	14.7935	-	-	-	-
A	14,834	14.7881	A v E	881	-0.0055	Dominated
D	15,976	14.7878	D v A	1,142	-0.0003	Dominated
B	16,801	14.7856	B v D	825	-0.0021	Dominated
C	17,972	14.7849	C v B	1,171	-0.0007	Dominated



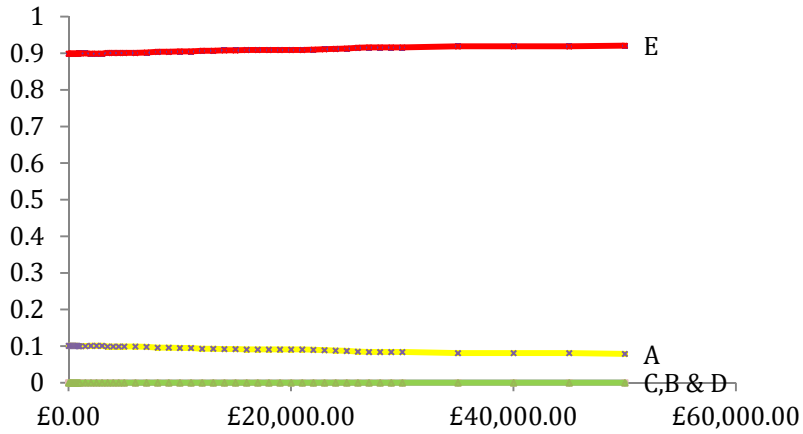
65a. CE Plane for 10-year time horizon



65c. CE Plane for lifetime horizon



65b. CEAC for 10-year time horizon



65d. CEAC for lifetime horizon

Figure 65. Cost-effectiveness acceptability curves for RS vs. THR age and gender adjusted using a bathtub model

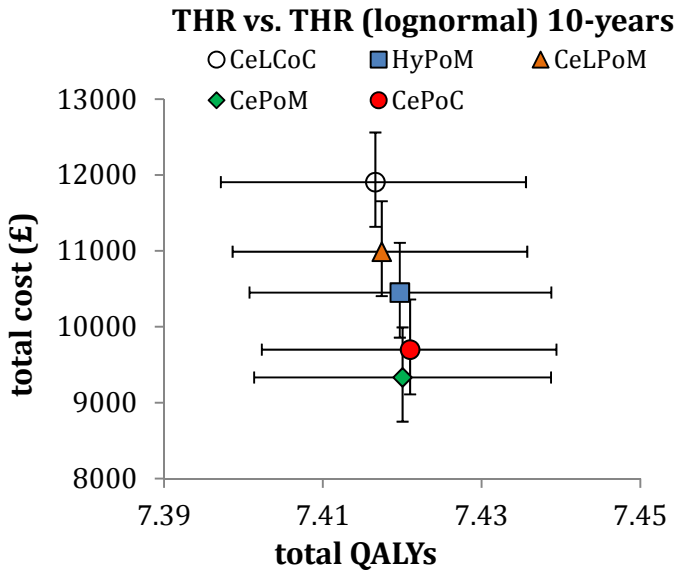
10.3.5 Sensitivity analyses: time to revision (lognormal model)

For this sensitivity analysis, we used a lognormal model time to revision to compare the cost-effectiveness of the different categories of THR. Table 108 shows that for both the deterministic and probabilistic analyses for both time horizons, Category A was cheaper, although Category E was more effective than the other four categories. The corresponding ICERs are reported in Table 108.

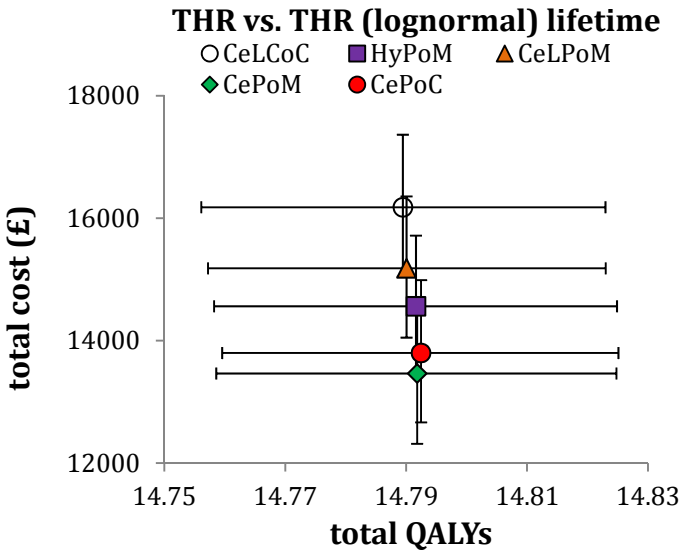
Table 108. Deterministic and probabilistic results for all THR patients using lognormal model

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,331	7.4203	-	-	-	-
E	9,690	7.4214	E v A	359	0.0010	342,781
D	10,446	7.4200	D v E	756	-0.0013	Dominated
B	10,986	7.4177	B v D	541	-0.0023	Dominated
C	11,901	7.4169	C v B	915	-0.0008	Dominated
Deterministic: Lifetime horizon						
A	13,476	14.7919	-	-	-	-
E	13,794	14.7926	E v A	318	0.0007	442,830
D	14,568	14.7917	D v E	773	-0.0009	Dominated
B	15,192	14.7901	B v D	624	-0.0016	Dominated
C	16,190	14.7895	C v B	998	-0.0006	Dominated
Probabilistic: 10-year time horizon						
A	9,334	7.4200	-	-	-	-
E	9,700	7.4210	E v A	366	0.0010	384,106
D	10,452	7.4197	D v E	752	-0.0013	Dominated
B	10,991	7.4174	B v D	539	-0.0023	Dominated
C	11,907	7.4166	C v B	916	-0.0008	Dominated
Probabilistic: Lifetime horizon						
A	13,464	14.7918	-	-	-	-
E	13,799	14.7924	E v A	335	0.0006	522,741
D	14,562	14.7916	D v E	762	-0.0008	Dominated
B	15,183	14.7900	B v D	621	-0.0016	Dominated
C	16,179	14.7894	C v B	997	-0.0006	Dominated

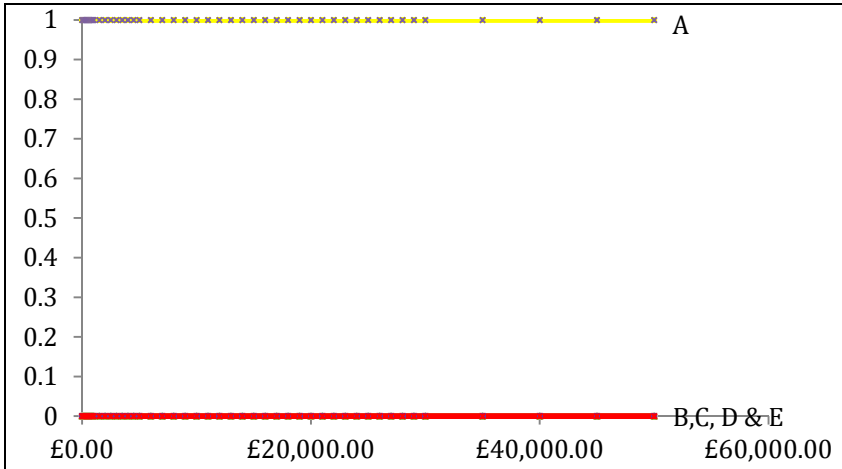
Figure 66 a and b show the cost-effectiveness planes with the 95% confidence intervals. For both the 10-year and lifetime horizons, although Category A is cheaper, Category E generates more QALYs. Figure 66 c and d shows the cost-effectiveness acceptability curves for THR vs. THR using a log normal model for the two time horizons. For both the 10-year time horizon and the lifetime horizon, if a decision maker was willing to pay anything from £0 to £50,000, Category A was nearly 100% cost-effective.



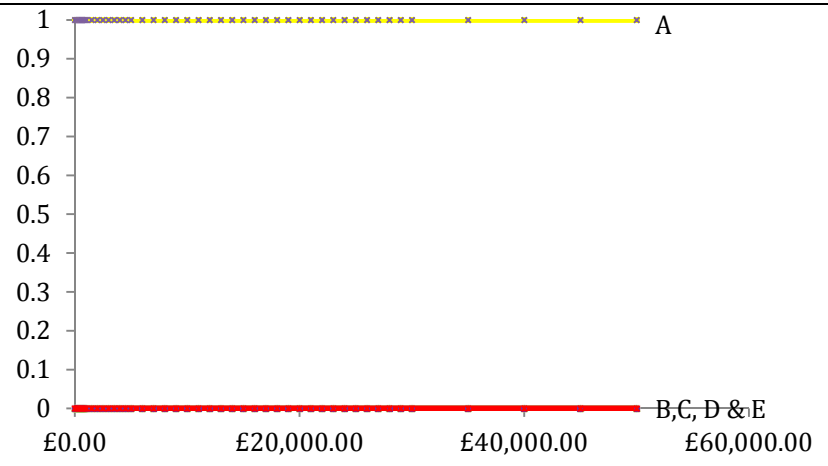
66a: CE plane for 10-year time horizon



66b: CE plane for lifetime horizon



66c: CEAC for 10-year time horizon



66d: CEAC for lifetime horizon

Figure 66. Cost-effectiveness planes and cost-effectiveness acceptability curves for THR vs. THR using a lognormal model

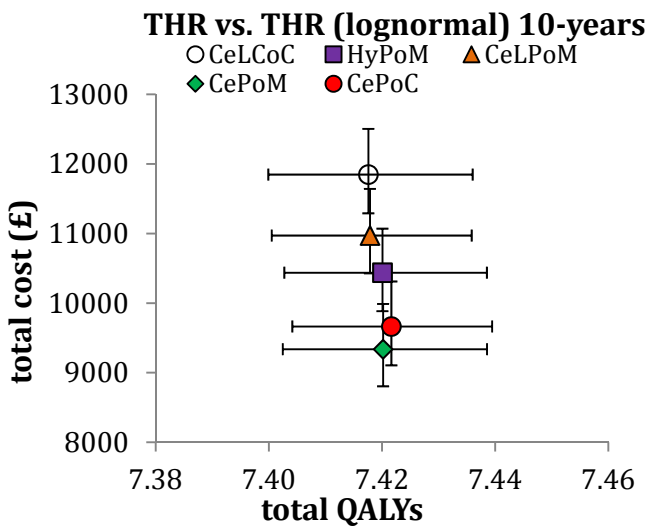
10.3.6 Sensitivity analyses: time to revision (lognormal model adjusted for age and gender)

For this sensitivity analysis, we used a lognormal model FOR time to revision adjusted for age and gender to compare the cost-effectiveness of the different categories of THR. Table 109 shows that for both the deterministic and probabilistic analyses for both time horizons, that Category A was cheaper; however, Category E was clearly more effective than the other four categories. The corresponding ICERs are also reported in Table 109.

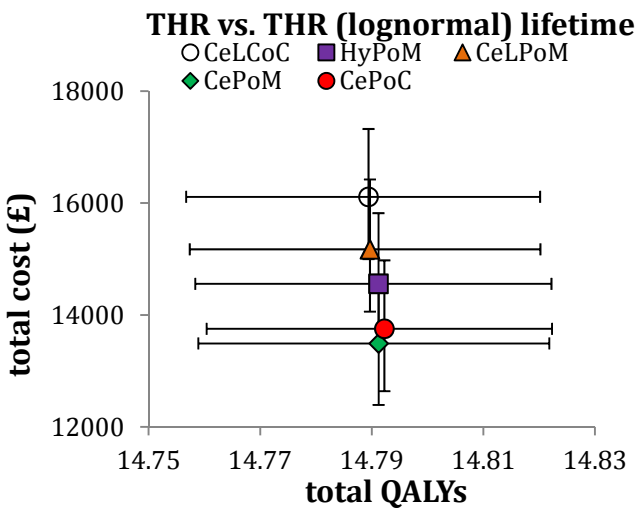
Table 109. Sensitivity analysis deterministic and probabilistic results for all THR patients - age and gender adjusted using a lognormal model

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,349	7.4201	-	-	-	-
E	9,667	7.4217	E v A	318	0.0016	202,741
D	10,446	7.4200	D v E	779	-0.0017	Dominated
B	10,982	7.4178	B v D	536	-0.0022	Dominated
C	11,858	7.4175	C v B	876	-0.0003	Dominated
Deterministic: Lifetime horizon						
A	13,505	14.7917	-	-	-	-
E	13,753	14.7928	E v A	248	0.0011	227,031
D	14,567	14.7917	D v E	814	-0.0011	Dominated
B	15,185	14.7902	B v D	618	-0.0015	Dominated
C	16,119	14.7899	C v B	934	-0.0002	Dominated
Probabilistic: 10-year time horizon						
A	9,339	7.4202	-	-	-	-
E	9,665	7.4216	E v A	327	0.0015	223,741
D	10,438	7.4201	D v E	773	-0.0016	Dominated
B	10,973	7.4179	B v D	534	-0.0022	Dominated
C	11,849	7.4176	C v B	877	-0.0003	Dominated
Probabilistic: Lifetime horizon						
A	13,493	14.7912	-	-	-	-
E	13,755	14.7923	E v A	263	0.0010	255,638
D	14,559	14.7912	D v E	804	-0.0011	Dominated
B	15,175	14.7897	B v D	616	-0.0015	Dominated
C	16,112	14.7894	C v B	937	-0.0003	Dominated

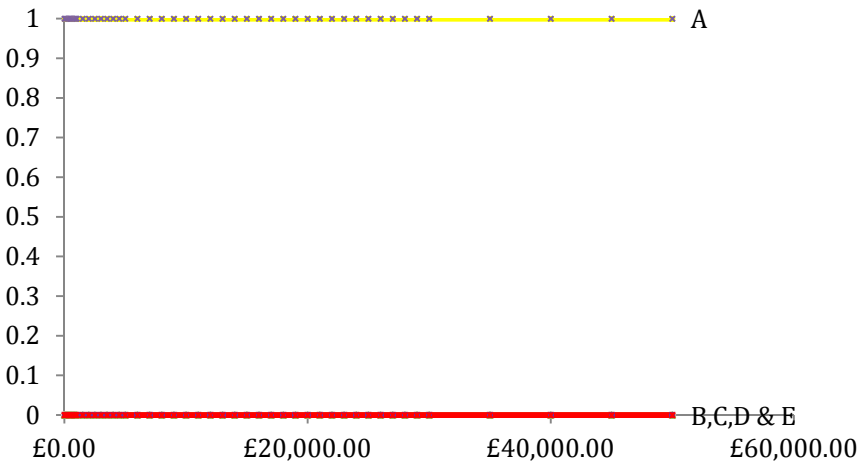
The corresponding cost-effectiveness planes are shown in Figure 67 a and c. For both the 10-year time horizon and the lifetime horizon, if the decision maker is willing to pay £20,000 per QALY, Category A is nearly 100% cost-effective (see Figure 67 b and d).



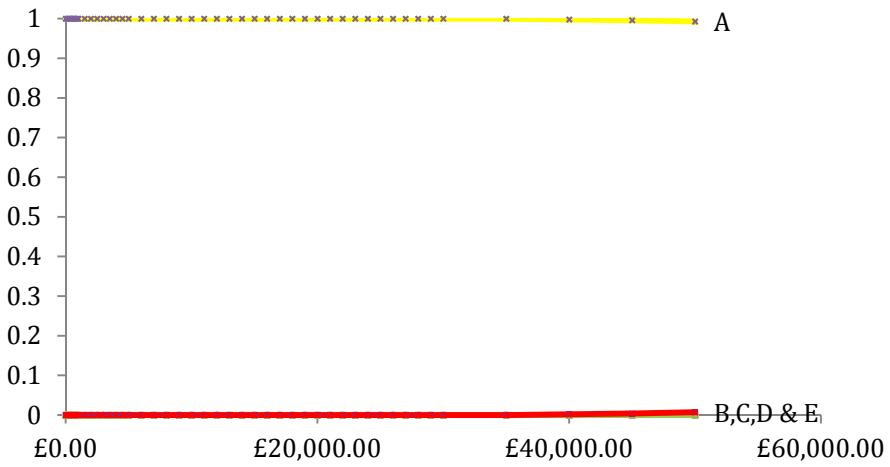
67a. CEAC for 10-year time horizon



67c. CE Plane using lognormal model age and gender adjusted



67b. CEAC for lifetime horizon



67d. CE Plane using lognormal model age and gender adjusted

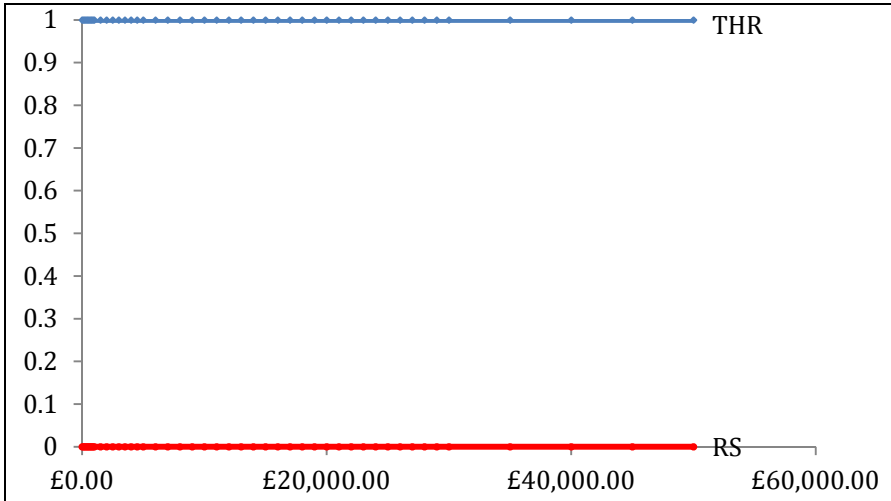
Figure 67. Cost-effectiveness acceptability curves for THR vs. THR age and gender adjusted using a lognormal model

10.3.7 Sensitivity analyses: costs

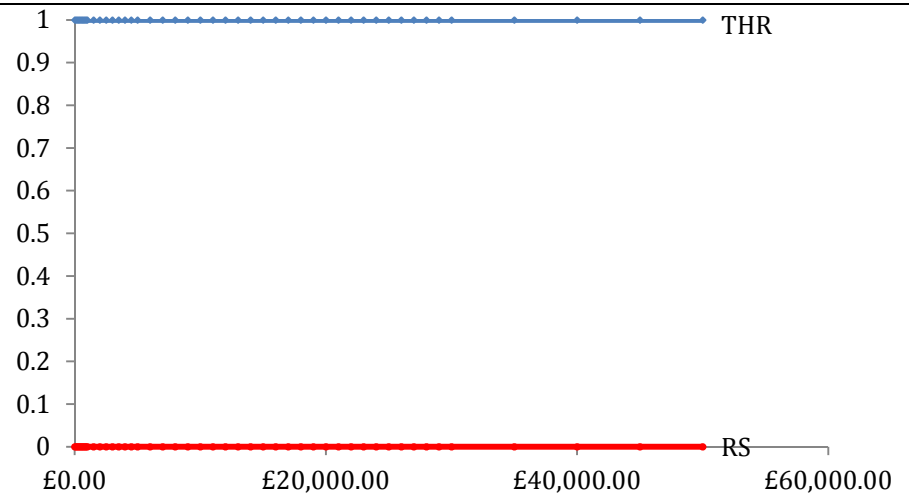
For this sensitivity analysis, we compared the cost-effectiveness of RS vs. THR using highest and lowest cost estimates for prostheses from the list prices supplied by the NHS supply chain. For both time horizons (10-year and lifetime) whether the lowest or highest costs were used, RS was dominated by THR (that is, RS was more expensive and less effective than THR). The corresponding CEACs are shown is Figure 68.

Table 110. Deterministic and probabilistic results for lowest and highest costs for THR vs. RS patients

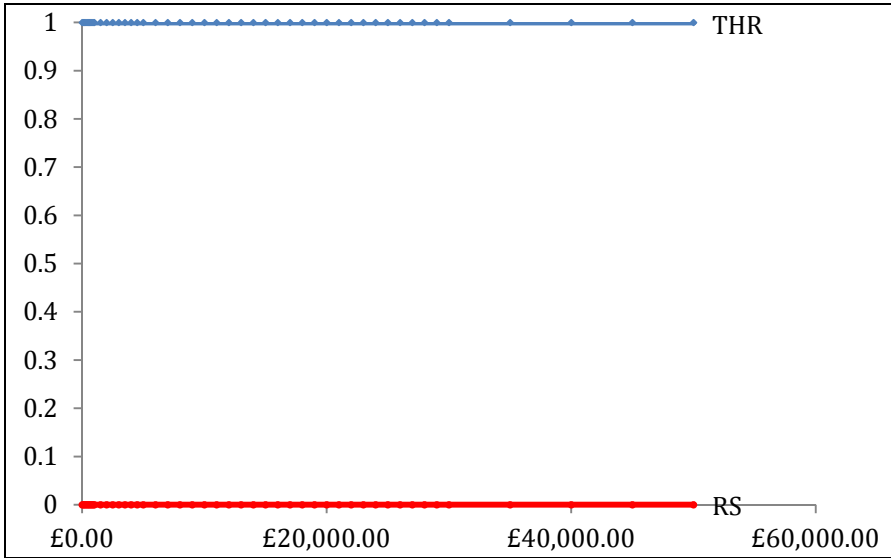
	Lowest cost		Highest cost	
	RS	**THR**	**RS**	**THR**
Deterministic: 10-year time horizon				
Total mean costs £	22,228	11,487	22,735	12,380
Total mean QALYs	7.2830	7.4147	7.2830	7.4147
Incremental cost £	10,741		10,355	
Incremental QALYs	-0.1317		-0.1317	
ICERs (£/QALY)	Dominated		Dominated	
Probabilistic: 10-year time horizon				
Total mean costs £	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	
ICERs (£/QALY)	Dominated		Dominated	
Deterministic: Lifetime horizon				
Total mean costs £	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	
ICERs(£/QALY)	Dominated		Dominated	
Probabilistic: Lifetime horizon				
Total mean costs £	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	
ICERs (£/QALY)	Dominated		Dominated	



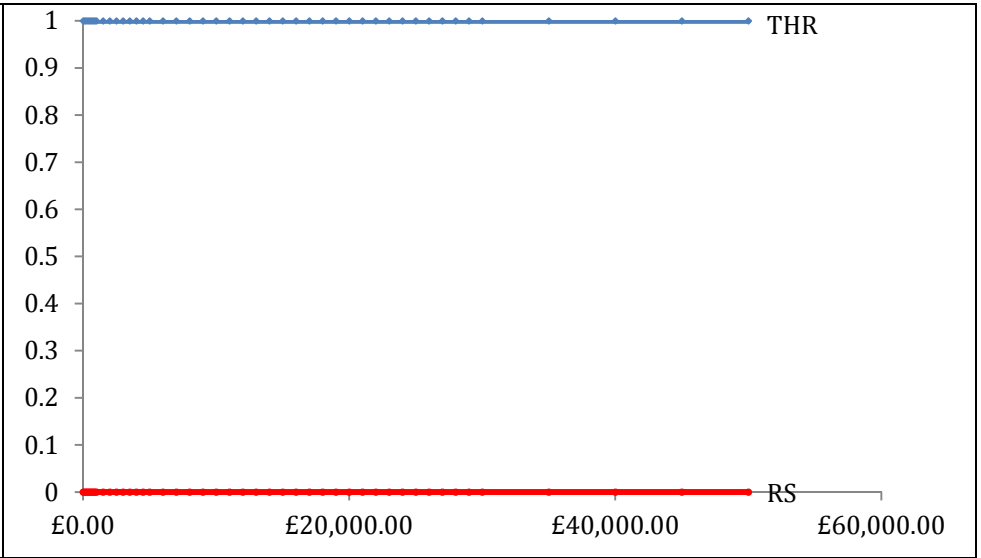
68a: CEAC for 10-year time horizon for lowest cost



68b: CEAC for lifetime horizon for lowest cost



68c: CEAC for 10-year time horizon for highest cost



68d: CEAC for lifetime horizon for highest cost

Figure 68. Cost-effectiveness acceptability curves for RS vs. THR using lowest and highest costs

For the THR vs. THR cost sensitivity analysis, 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 111 shows 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 the ICER was £297,098 per QALY gained. For the lifetime horizon, Category E dominated the other four categories. The corresponding CEACs are shown in Figure 69.

Table 111. Deterministic and probabilistic results using the highest prices for all THR patients using a bathtub model

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,675	7.4189	-	-	-	-
E	10,018	7.4207	E v A	343	0.0018	190,326
D	10,918	7.4182	D v E	900	-0.0025	Dominated
B	11,913	7.4156	B v D	995	-0.0026	Dominated
C	12,977	7.4143	C v B	1,064	-0.0013	Dominated
Deterministic: Lifetime horizon						
E	14,798	14.7909	-	-	-	-
A	15,032	14.7887	A v E	235	-0.0022	Dominated
D	16,371	14.7881	D v A	1,338	-0.0006	Dominated
B	17,562	14.7861	B v D	1,192	-0.0020	Dominated
C	19,091	14.7845	C v B	1,529	-0.0016	Dominated
Probabilistic: 10-year time horizon						
A	9,672	7.4191	-	-	-	-
E	10,055	7.4204	E v A	383	0.0013	297,098
D	10,917	7.4184	D v E	862	-0.0020	Dominated
B	11,909	7.4158	B v D	992	-0.0026	Dominated
C	12,973	7.4145	C v B	1,063	-0.0013	Dominated
Probabilistic: Lifetime horizon						
E	14,814	14.7909	-	-	-	-
A	15,030	14.7889	A v E	217	-0.0020	Dominated
D	16,378	14.7883	D v A	1,347	-0.0007	Dominated
B	17,570	14.7863	B v D	1,193	-0.0020	Dominated
C	19,076	14.7848	C v B	1,506	-0.0015	Dominated

Using the lowest cost estimates for prostheses from the list prices supplied by the NHS supply chain, Table 112 shows 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 69.

Table 112. Deterministic and probabilistic results using the lowest prices for all THR patients using a bathtub model

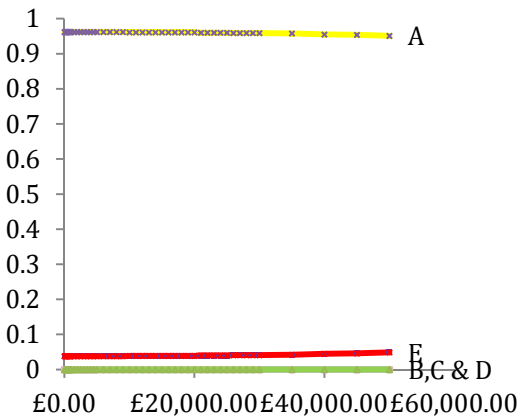
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,046	7.4189	-	-	-	-
E	9,322	7.4207	E v A	277	0.0018	153,663
D	10,080	7.4182	D v E	758	-0.0025	Dominated
B	10,801	7.4156	B v D	721	-0.0026	Dominated
C	11,750	7.4143	C v B	949	-0.0013	Dominated
Deterministic: Lifetime horizon						
E	14,102	14.7909	-	-	-	-
A	14,402	14.7887	A v E	301	-0.0022	Dominated
D	15,533	14.7881	D v A	1,130	-0.0006	Dominated
B	16,450	14.7861	B v D	918	-0.0020	Dominated
C	17,864	14.7845	C v B	1,414	-0.0016	Dominated
Probabilistic: 10-year time horizon						
A	9,042	7.4187	-	-	-	-
E	9,326	7.4204	E v A	283	0.0017	165,912
D	10,081	7.4180	D v E	755	-0.0024	Dominated
B	10,799	7.4154	B v D	719	-0.0026	Dominated
C	11,750	7.4140	C v B	950	-0.0013	Dominated
Probabilistic: Lifetime horizon						
E	13,618	14.7917	-	-	-	-
A	14,391	14.7887	A v E	773	-0.0040	Dominated
D	15,534	14.7870	D v A	1,143	-0.0007	Dominated
B	16,437	14.7851	B v D	903	-0.0020	Dominated
C	17,840	14.7835	C v B	1,403	-0.0016	Dominated

For this sensitivity analysis, we compared the cost-effectiveness for all THR patients using a 20% price de-escalator to reflect in reality what NHS trusts would pay for the implants.

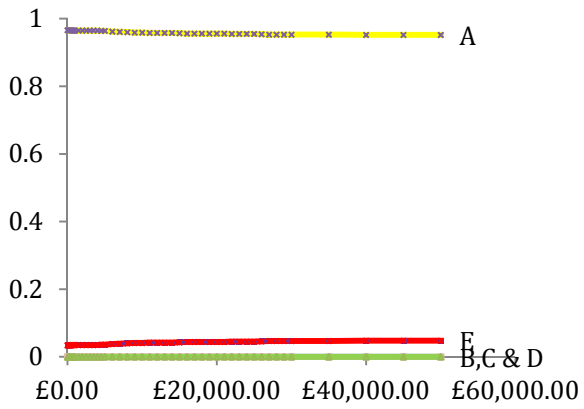
Table 113 shows for the 10-year time horizon, although Category A was cheaper, Category E was more effective. For the lifetime horizon, Category E dominated the other four categories. The corresponding CEACs are shown in Figure 69.

Table 113. Deterministic and probabilistic results assuming a price de-escalator of 20% for all THR patients using a bathtub model

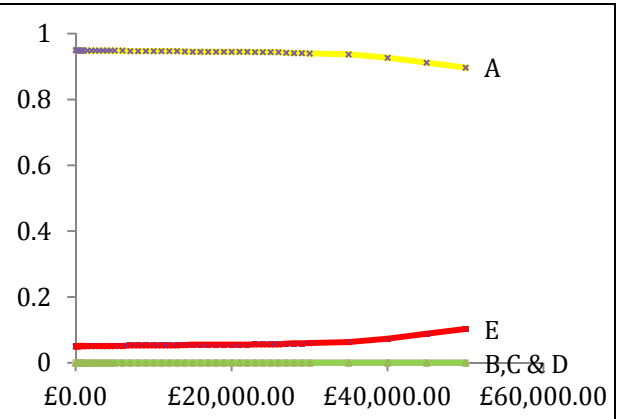
Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,132	7.4189	-	-	-	-
E	9,344	7.4207	E v A	212	0.0018	117,489
D	10,058	7.4182	D v E	714	-0.0025	Dominated
B	10,552	7.4156	B v D	494	-0.0026	Dominated
C	11,338	7.4143	C v B	786	-0.0013	Dominated
Deterministic: Lifetime horizon						
E	14,123	14.7909	-	-	-	-
A	14,489	14.7887	A v E	366	-0.0022	Dominated
D	15,510	14.7881	D v A	1,021	-0.0006	Dominated
B	16,201	14.7861	B v D	690	-0.0020	Dominated
C	17,452	14.7845	C v B	1,252	-0.0016	Dominated
Probabilistic: 10-year time horizon						
A	9,138	7.4184	-	-	-	-
E	9,296	7.4209	E v A	158	0.0025	62,906
D	10,066	7.4177	D v E	770	-0.0032	Dominated
B	10,558	7.4155	B v D	492	-0.0026	Dominated
C	11,342	7.4138	C v B	784	-0.0013	Dominated
Probabilistic: Lifetime horizon						
E	14,012	14.7910	-	-	-	-
A	14,484	14.7883	A v E	472	-0.0026	Dominated
D	15,504	14.7877	D v A	1,020	-0.0006	Dominated
B	16,193	14.7857	B v D	689	-0.0020	Dominated
C	17,450	14.7841	C v B	1,257	-0.0016	Dominated



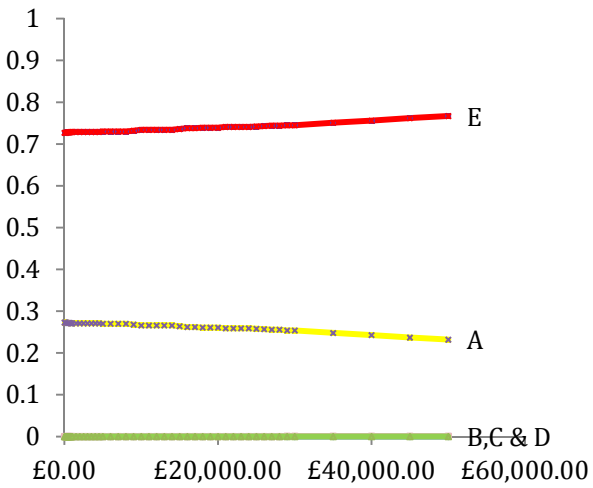
69a: CEAC for 10-year time horizon for lowest cost



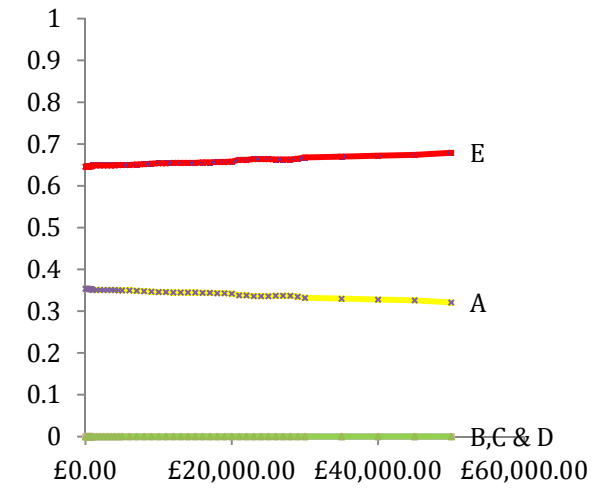
69b: CEAC for 10-year time horizon for highest cost



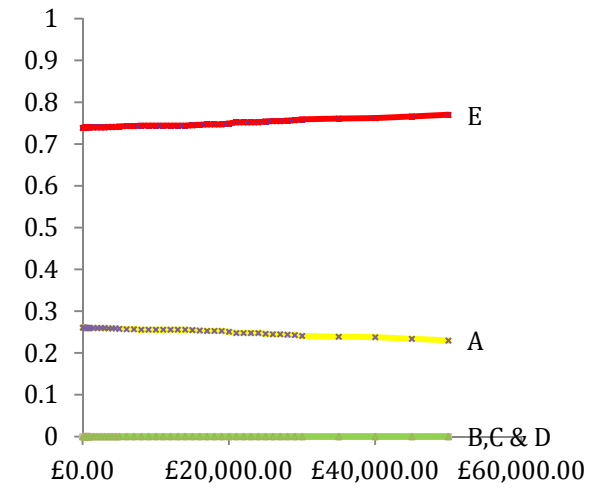
69c: CEAC for 10-year time horizon for 20% de-escalator



69d: CEAC for lifetime horizon for lowest cost



69e: CEAC for lifetime horizon for highest cost



69f: CEAC for life- time horizon for 20% de-escalator

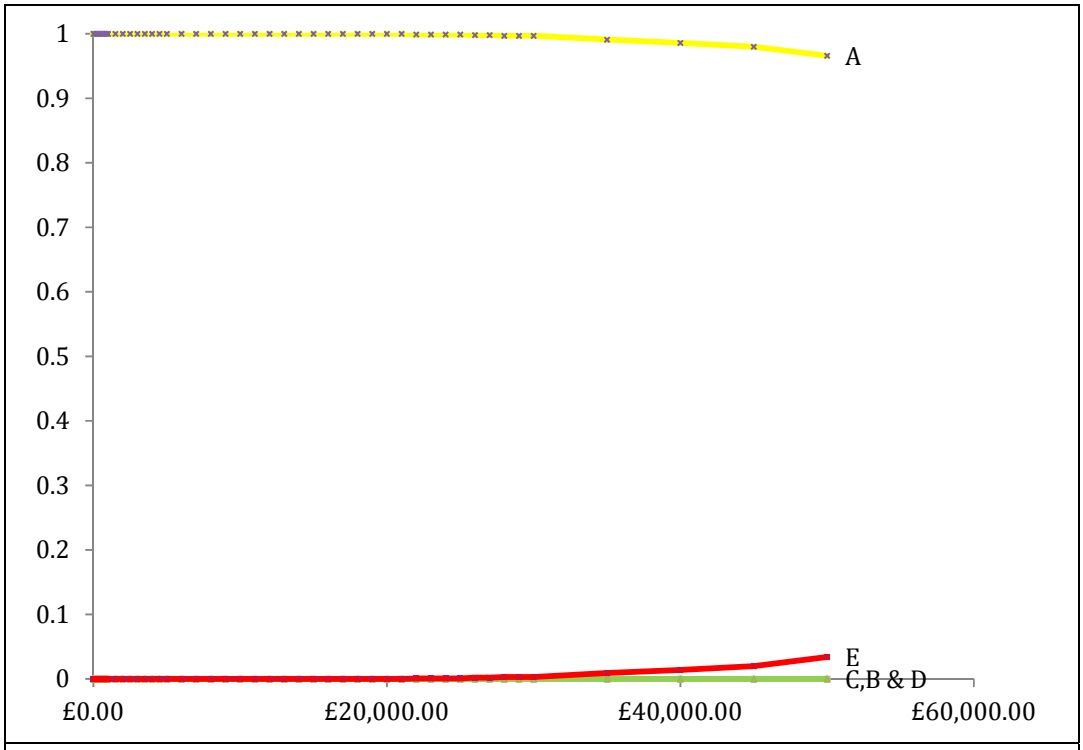
Figure 69. Cost-effectiveness acceptability curves for all THR patients using lowest and highest costs and assuming a 20% price de-escalator

10.3.8 Sensitivity analyses: utilities

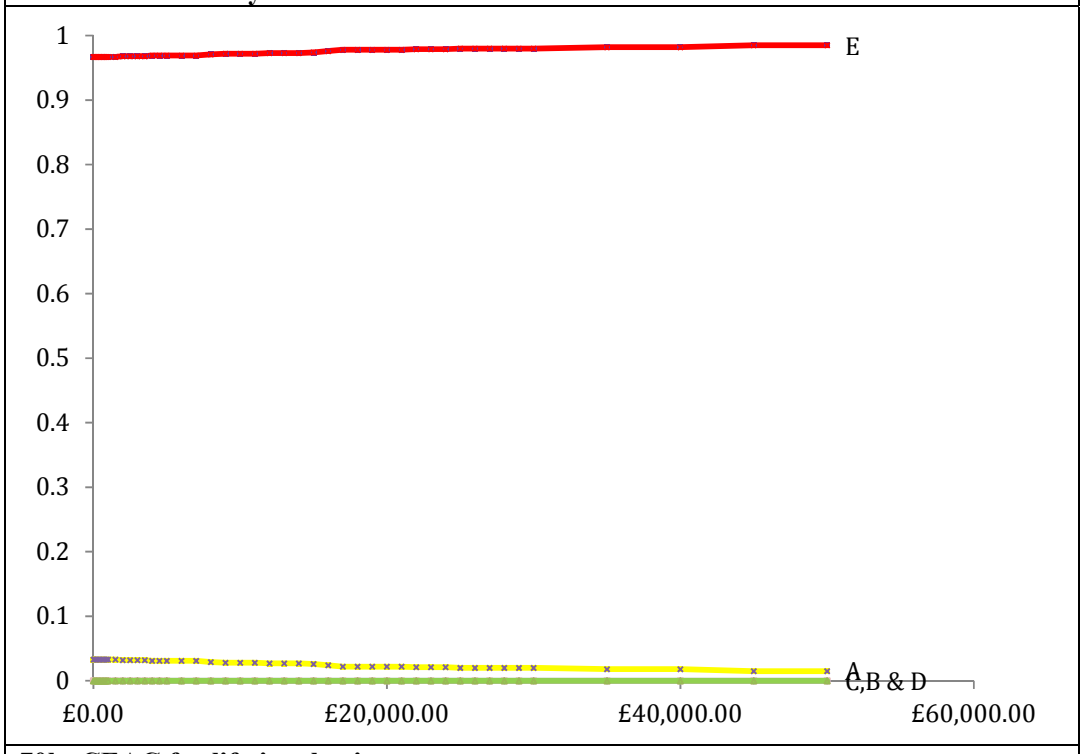
For this sensitivity analysis, utility values from [297] were used in the Markov model. Table 114 shows the deterministic and probabilistic results for the 10-year and lifetime horizons. For the 10-year time horizon (both deterministic and probabilistic) Category A was cheaper than all four categories; however, slightly more QALYs were generated for Category E than the other four categories. The incremental cost-effectiveness ratio between Category A and Category E was £153,067 per QALY gained for the deterministic analysis and was £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 70.

Table 114. Deterministic and probabilistic using utility values from Rolfson

Category	Total mean costs £	Total mean QALYs	Comparison	Incremental costs £	Incremental QALYs	ICER (£)
Deterministic: 10-year time horizon						
A	9,444	7.5764	-	-	-	-
E	9,743	7.5783	E v A	299	0.0020	153,067
D	10,588	7.5757	D v E	845	-0.0027	Dominated
B	11,155	7.5728	B v D	567	-0.0029	Dominated
C	12,112	7.5714	C v B	957	-0.0014	Dominated
Deterministic: Lifetime horizon						
E	14,522	15.1174	-	-	-	-
A	14,801	15.1146	A v E	278	-0.0028	Dominated
D	16,040	15.1139	D v A	1,240	-0.0007	Dominated
B	16,804	15.1115	B v D	764	-0.0024	Dominated
C	18,226	15.1094	C v B	1,422	-0.0021	Dominated
Probabilistic: 10-year time horizon						
A	9,443	7.5760	-	-	-	-
E	9,741	7.5780	E v A	298	0.0020	150,644
D	10,590	7.5752	D v E	848	-0.0027	Dominated
B	11,153	7.5724	B v D	564	-0.0028	Dominated
C	12,114	7.5709	C v B	960	-0.0015	Dominated
Probabilistic: Lifetime horizon						
E	14,504	15.1178	-	-	-	-
A	14,795	15.1149	A v E	291	-0.0029	Dominated
D	16,023	15.1142	D v A	1,228	-0.0007	Dominated
B	16,807	15.1118	B v D	784	-0.0024	Dominated
C	18,208	15.1098	C v B	1,402	-0.0020	Dominated



70a: CEAC for 10-year time horizon



70b: CEAC for lifetime horizon

Figure 70. Cost-effectiveness acceptability curves for all THR using Rolfson et al (2011) utility values

10.3.8.1 One-way sensitivity analysis - Tornado diagram

We undertook sensitivity analysis where we varied a number of important variables whilst holding others constant in order to compare the relative importance of particular variables in driving our estimates of lifetime net monetary benefit of CePoC (category E) vs. CePoM (category A) at a willingness to pay threshold of £20,000. The tornado diagram (Figure 71) illustrates our findings. For each variable the diagram indicates the changes to the inputs.

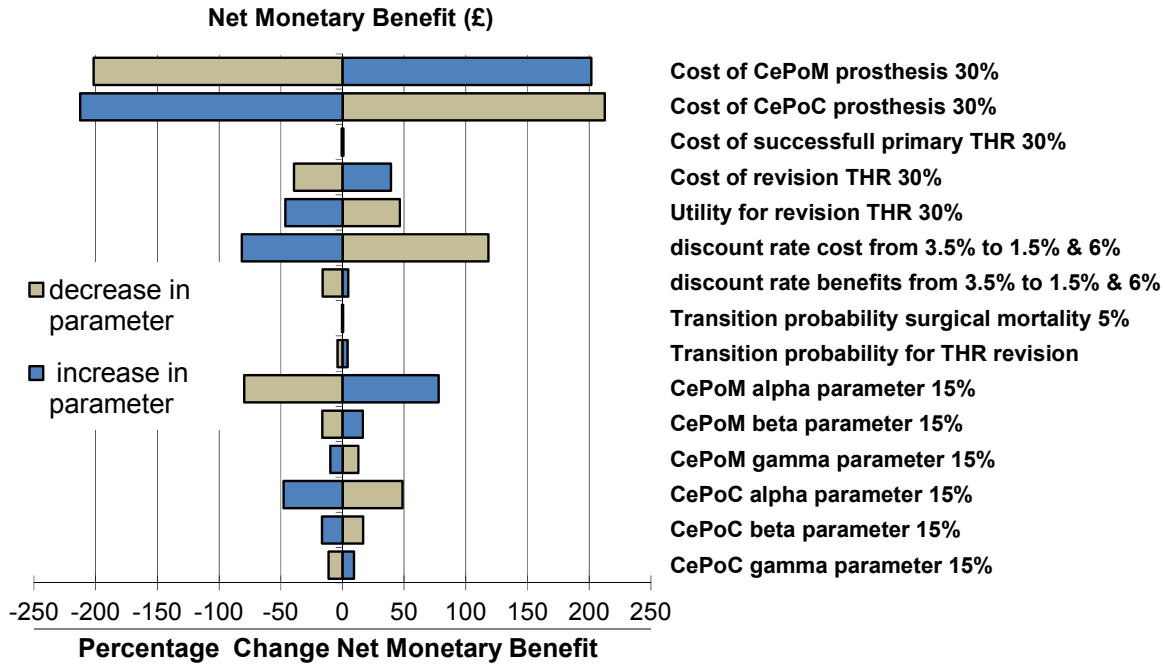


Figure 71. Tornado diagram illustrating sensitivity analysis for net monetary benefit: CePoP versus CePoM

The diagram is centred around net monetary benefit of category E (CePoC) versus category A (CePoM) at a willingness to pay threshold of £20,000 (£321). We can see that the cost of the prosthesis is the most important factor, and that for each of CePoC and CePoM 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 also is the CePoM alpha parameter – that is the revision rate setting for CePoM within the model.

10.3.9 Discussion of economic assessment

We built a Markov, multi state model to investigate both RS and THR. Health states included successful primary surgery, revision surgery, successful revision surgery and death. Cycle length was one 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 pounds sterling (£) at 2011/2012 prices. Health outcomes were measured in quality-adjusted life years. 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 1,000 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 gender-adjusted utility values from the PROMs dataset for both THR and RS. For the comparison of RS versus THR we undertook sensitivity analyses stratified by gender 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 cost-effectiveness acceptability curves comparing RS with THR overall and in separate age groups at different levels of willingness to pay.

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 where appropriate. We used highest and lowest list prices supplied by the NHS supply chain in sensitivity analyses. We used age and gender-adjusted utility values from the PROMs dataset for before and after hip replacement and for revision.

We undertook sensitivity analyses and analysis of cost drivers including investigating age and gender categories, stratifying by age (less than and more than 65 years), different methods of extrapolation of revision rates (using a lognormal model), and by varying prosthesis costs (using NHS list prices) and discount rates. We constructed cost-effectiveness acceptability curves comparing different types of THR overall and in separate age groups at different levels of willingness to pay.

10.3.10 Summary of results

We found that revision rates for all RS, compared to THR (all THR, all of our identified categories of THR combined, each of our THR categories separately) were always higher.

The weighted mean cost of the THR prostheses obtained from the NHS supply chain was £2,571. The prosthesis cost for RS was sourced from the NHS Supply Chain and was reported as (£2,672) £101 more than the cost of THR. This corresponds with the literature where the cost of RS has been reported as more expensive than THR.[40] For all analyses, mean costs for RS were higher than THR; and 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 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 analysed separately in sensitivity analyses for males and females by age group (40, 50 and 60 years). For all age and gender groups RS remained clearly dominated by THR. CEACs showed that for all patients, THR was almost 100% cost-effective at any willingness to pay level.

For different types of total hip replacement, 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: CePoM (Cemented-cemented with a polyethylene-metal articulation) (125,285 patients); Category B: CeLPoM (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: HyPoM (Hybrid (cementless-cemented) with a polyethylene-metal articulation) (28,471 patients) and Category E: CePoC (Cemented-cemented with a polyethylene-ceramic articulation) (12,075 patients).

There were age and gender differences in the populations with different types of THR and variations in revision rates. For all interventions, revision rates at nine years were substantially less than the benchmark of 10% (Category A: 2.5%; B: 3.2%; C:3.5%; D:2.5%; E:1.6 at 9 years). Costs of the different prostheses were as follows: Category A – CePoM £1,557.38; B – CeLPoM £3,015.60; C – CeLCoC £3,868.80; D – HyPoM £2,649.78; E – CePoC £1,995.98.

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

For example in the deterministic analysis, compared to Category E, Category A (CePoM) cost £278 more (£14,801 compared to £14,523) and generated 0.0022 fewer QALYs (14.7887 as compared to 14.7909) and 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 to Category A (2.8%) at a willingness to pay of £20,000 per QALY. For patients aged over 65 years, at a willingness to pay of £20,000 per QALY, Category A was 100 % cost-effective.

Sensitivity analyses using a lognormal model (indicating a decreasing risk of revision over time) for extrapolation beyond the observed data for revision rates, found category A to be cheaper at a lifetime horizon for all age-gender groups combined. Although category E was more effective than the other four categories, Category A was 100% cost-effective at a willingness to pay threshold of £20,000 per QALY. Further sensitivity analysis using an age and gender- adjusted log normal model demonstrated the same finding: that at a lifetime horizon and a willingness to pay of £20,000 per QALY, Categories A was 100% cost-effective.

Using a one-way sensitivity analysis and varying the main inputs (e.g., costs by 30%) in the base case analysis for all age-gender groups, and comparing Category A with Category E, demonstrated that the main drivers of difference were costs of components, discount rates and modelled revision rates.

10.3.10.1 *Strengths and limitations*

Although we undertook a rigorous systematic review for cost-effectiveness studies, we could only identify one cost-utility analysis of RS versus THR. The study reported NHS and PSS cost for the 12 months post hip replacement.[40] The costs for a successful primary procedure were taken from the literature. Although the figures included all costs relevant to in-hospital stay, they do not include costs of long term follow-up post-discharge (after 12 months). Therefore, the cost of follow-up was taken from Edlin et al (2012).[40] We assumed the cost of follow-up to be the same for the first year and for all other 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 cost and resource use of adverse events other than those requiring revision.

The marginal difference in QALYS is negligible between Categories A to E. On the basis of a negligible difference in QALYs, it is therefore difficult to make a fair comparison between them in terms of outcomes. However, costs of the prostheses vary. Category A was less expensive when compared to category E and in the base case category E generated more QALYs over a lifetime horizon. The prices for prostheses were obtained from the NHS supply chain and reflect list prices in line with the NICE reference case.[365] 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. At a lifetime horizon, category E was less costly and more effective. This sensitivity analysis found that the category E remained cost-effective even with changes to the prosthesis cost.

The cost of prostheses varied depending on which category was used in 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 over-estimated actual revision prosthesis costs but reflected a fair comparison across groups.

We tested whether our results were robust to alternative time to revision models. In the base case analysis the revision rates were modelled using a bathtub model where a high hazard for failure associated with surgery is followed by a decreasing hazard that plateaus during initial recovery period, and is then followed by 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 lognormal model – a decreasing hazard model. Using this scenario, Category A was less costly and less effective and category E was more costly and more effective at both 10-year and lifetime horizons. The decreasing hazard model is unlikely to capture increasing likelihood of revision due to wear and tear in younger age group. Hence, we undertook another sensitivity analysis where we modelled revision rates based on both bathtub and lognormal fits but adjusted for age and gender.

The utilities for the revision health state were based on PROMS data, however, PROMS data do not discriminate different types of further surgery, so some utilities reported might be reflecting interventions other than revision. However, because in our model revision rate differences only affect utility for one year, the impact of revision rates on the overall QALYs is minimal. We were unable to incorporate adverse events which 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 for linking these. However, we were able to undertake sensitivity analyses to take account of possible costs 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 the structure in relation to these external comments.

10.3.11 Conclusion of cost-effectiveness analysis

Compared to THR, revision rates for RS were higher, mean costs for RS were higher and mean QALYs gained were lower; RS was dominated by THR. Very similar results were obtained for deterministic and probabilistic results and for all age and gender groups and THR was almost 100% cost-effective at any willingness to pay level.

Revision rates for all types of THR were low. Costs of prostheses varied depending partly on complexity (e.g. presence or absence of a liner etc.). There were small but clear differences between categories in both costs and effectiveness as measured by QALYs and when age and gender groups were factored in. The mean total cost for Category A was slightly lower and mean QALY gain for Category E was slightly higher for older age groups where revision rates are lower. However, across all age gender groups combined, for the base-case analysis, mean costs for Category E (CePoC) were slightly lower and 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.

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 spread across 1,000s of iterations.

10.4 Comparison of results with TA2, TA44, Manufacturer's submission, & international registries

NICE guidance TA2 (April 2000) suggests a benchmark revision rate of 10% or less at 10 years. Similarly TA 44 (June 2002) suggests this bench mark, or a three year equivalent for resurfacing.[355] The available evidence underpinning the benchmark is old and incomplete relative to that currently available in the UK NJR and other registries. While the THR prostheses examined in this report easily satisfy this benchmark the requirement for revision after RS did not (see Chapter 8).

One manufacturer, DePuy, submitted a review and economic analysis. Analyses of the following interventions were presented: cemented THR, cementless THR, Hybrid THR, reverse Hybrid THR and resurfacing. Except for resurfacing, 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 requirement for revisions were broadly similar to those reported in Chapter 8. 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 gender and generated a monotonically decreasing hazard with time. The manufacturer justified the choice of decreasing hazard "*because all previous economic evaluations which assumed parametric distributions assumed Weibull distributions"*.

This statement is misleading since each of the economic evaluations referenced in fact employed two rather than one Weibull model, one to early and one to late revisions, so that the resulting hazard followed a "U" shaped bath tub function and not a monotonic function with decreasing hazard as used by the manufacturer. The manufacturer's models predicted decreasing hazard on extrapolation beyond observed data but the requirements for revision beyond 10 years were not tabulated. Therefore because of this lack of accessible data and since different prostheses were analysed, any comparison with the present results is problematic and unlikely to be informative. Two major registries, the Swedish and Australian, provide longer term follow up of patients than the NJR from which reliable data to only about nine years is available.

These registries consider smaller numbers of patients but the Swedish registry provides relevant information for 19 years follow up. The bath tub model of hazard for revision infers that revision rates will gradually increase at some time after plateauing and this is supported by data in both these registries. Figure 72 shows time to revision for different age groups reported from the Swedish registry. This shows increasing rates of revision from between about five and 15 years of follow up for most age groups; for these this data is consistent with a bath tub hazard. For the oldest age group revision rates are relatively low and are probably not consistent with the bath tub model. Similar results are found from the Australian registry.

This section aimed to compare the results of Warwick economic model with TA2, TA44, and manufacturer's model. However, it must be noted that as we do not cover the same comparators we cannot directly compare models and findings.

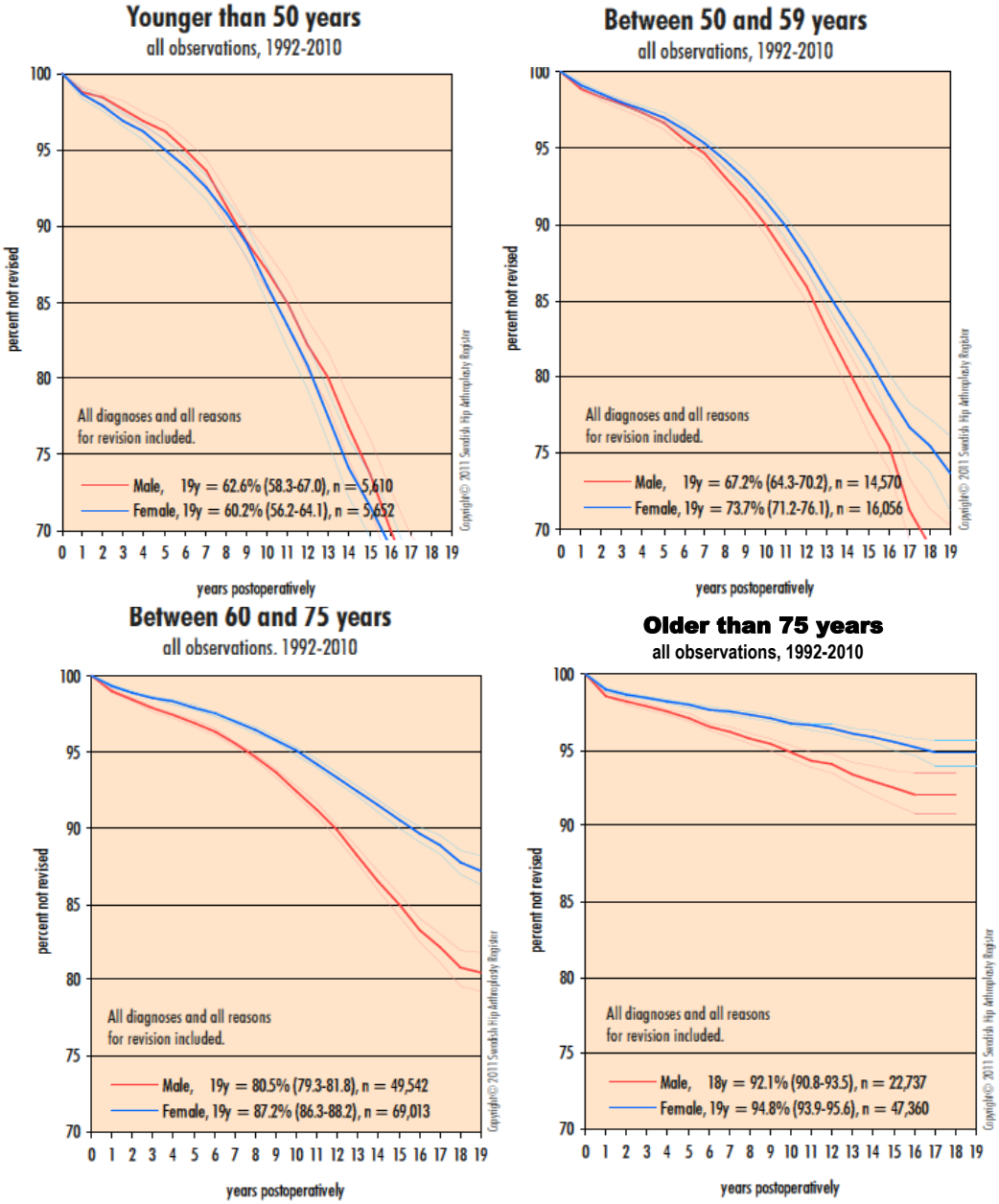


Figure 72. Swedish Registry data for time to revision to 19 years of follow up

It should be borne in mind that long follow up (e.g., up to 20 years) necessitates looking at devices and practices that now may no longer be widely used. The NJR data provided observed rates up to between 9 and 10 years only but may better reflect modern practice.

Further support for a bath tub model comes from the RCT by Kim et al. (2011)[126] who reported extended follow up to about 20 years; the reported revision rates were higher between 15 and

twenty years than between 10 and 15 years. Several long follow up observational studies provide similar evidence as illustrated in Figure 73.

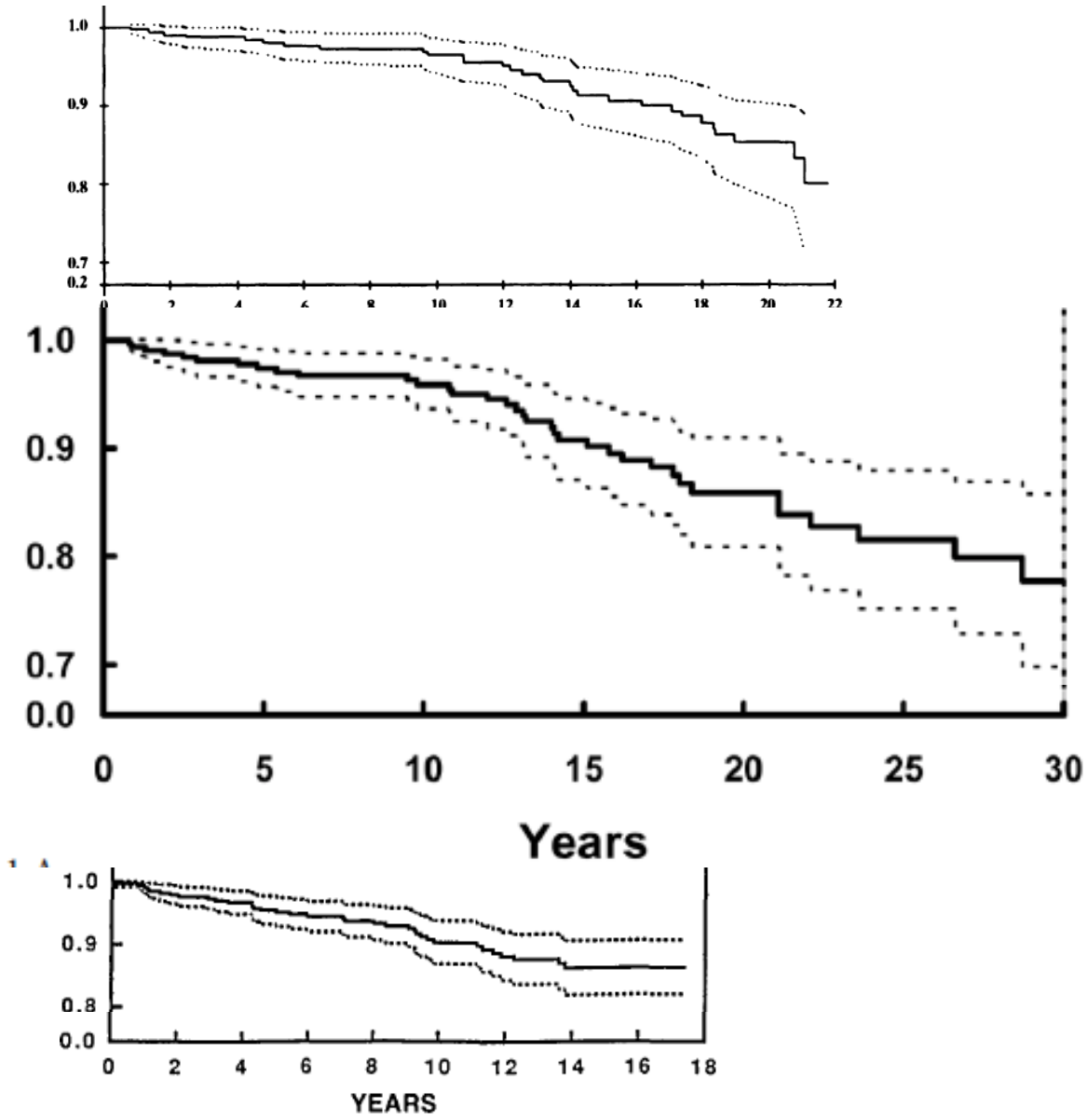


Figure 73. Time to revision results for Schulte, Madey, and Callaghan, respectively (top to bottom graphs) Shulte 1993 , Madey 1997, Callaghan 2000.

10.5 Summary and critique of manufacturer's submissions

Four manufacturer submissions were received (DePuy International Ltd., Smith & Nephew, Inc., Stryker, JRI Orthopaedics Limited). The following section provides 1) a description of the submissions, 2) an evaluation of the literature searches, 3) the limitations and strengths of clinical effectiveness reviews, 4) the overall quality considerations for the cost-effectiveness reviews, 5) a critique of the model structure (if possible), and 6) the main conclusions identified by the review team for each submission.

10.5.1 DePuy International Ltd

Contents of submission

DePuy provided an economic model in Excel and a 244 page technology assessment of the clinical 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 total hip replacement and hip resurfacing arthroplasty compared with each other for people in whom both procedures are suitable
- Different types of primary total hip replacement compared with each other for people in whom hip resurfacing arthroplasty 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 Markov model. The model considered the following hip replacement procedures:

- Cemented THR
- Cementless THR
- Hybrid THR
- Reverse hybrid THR
- Hip resurfacing arthroplasty

Data for the model was generally derived from the English National Joint Register (revision rates), the literature (utility data) and a micro-costing analysis (costs). The National Patient Reported Outcome Measure database and the New Zealand Joint Registry were further data sources.

The overall conclusions were that THR dominated hip resurfacing arthroplasty in patients suitable for both procedures and [redacted] was the optimal treatment strategy for patients, both suitable and unsuitable for hip resurfacing. 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 be based not solely on 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 manufacturer submission are thorough and accurate. However, there are several concerns; 1) that the Medline In-Process 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 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 following systematic review included in the current report would not have been retrieved, because it has not been indexed for age (see Ethgen et al., 2004).[188] 3) A grey literature search was not undertaken.

Strengths and limitations of clinical effectiveness review

a) *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, comprises a comprehensive search and pre-specified the inclusion/exclusion criteria. The screening of identified evidence and data extraction of eligible studies was carried out independently and the study and baseline population characteristics are well presented in tables.

b) 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 if 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, these 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 due to clinical heterogeneity, inconsistent results, etc. 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/QALY outcomes; this might have restricted the review missing studies reporting basic costs and/or resource use for patients undergoing THR or resurfacing. The study selection is transparent. However, no table of excluded studies with reason is given. The review did not provide a standardised quality assessment of the included studies nor for 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 one key study was selected to provide data on utilities and costs for the model, respectively. These two studies,[38,297] investigated THR only and did not sufficiently 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 2010. The structure of the model is consistent with previous cost-effectiveness models on total hip replacement 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 for which hip RS or THR are suitable; and for patients who are not suitable for RS, received THR. The population selected and the interventions and comparators are appropriate, as outlined in the NICE scope. The model assumes a quarter-year cycle and a life-time 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 THR

The categories of THR included in the model by DePuy comprised of cemented, cementless, hybrid, reverse-hybrid DePuy [redacted] and DePuy [redacted]. No clear justification was given that explained the choice of the categories.

Methods defining the population for whom resurfacing is suitable / unsuitable

The definition of the two populations of patients, those for whom both THR and hip resurfacing arthroplasty are suitable, and those for whom THR is suitable but hip resurfacing arthroplasty is not suitable, was based on patient level data provided by the NJR. DePuy assumed that "…*the population characteristics of patients suitable for hip resurfacing arthroplasty or THR in the base case were those of patients that received hip resurfacing arthroplasty procedures that were recorded in the NJR*" and that "*the population characteristics of patients not suitable for hip resurfacing arthroplasty were assumed to be the same as patients who received THR procedures that were recorded in the NJR …*" (page 101 DePuy submission). The population characteristics of the two population groups are given in Table 115. The impact of this assumption was tested in subsequent sensitivity analyses.

Table 115. Patient populations considered in the manufacturer's model based on NJR data

Population	Mean age (years)	Male, %	Assumption
Hip resurfacing	55.3	70.9	Suitable for THR and resurfacing arthroplasty
THR	70.4	37.5	Not suitable for resurfacing arthroplasty
All patients	69.2	40.5	

Resource costs and utility values used in the DePuy model

Table 116 and

Table 117 list the utilities and costs used in the model by DePuy, sources of the values and the manufacturer's justification for using the values.

Table 116. Utility values used in DePuy model

	Value	**Source**	**Justification by Manufacturer**
Pre-operative utility	0.41	Rolfson et al. (2011)	This study had a very large sample size (32,396 patients from the Swedish Hip Arthroplasty Register) and reported pre-operative and one year post-operative utility values. Reported EQ-5D scores using the UK EQ-5D tariff (page 103 DePuy submission)
Post-operative utility	0.78	Rolfson et al. (2011)	
Post-revision disutility	0.145	Dawson et al. (2001)	To reflect the lower QoL associated with a subsequent surgical intervention which was considered appropriate by clinical experts (page 115 DePuy submission)

Table 117. Values for resource costs used in DePuy model

	Value	Source	Justification by Manufacturer
Prosthesis costs (page 118 DePuy submission)			
Cemented	£1,029.00	Unit costs: DePuy list prices No. units: Assumption based on NJR 9th Annual Report 2012	List price prosthesis costs in line with the NICE reference case (page 116 DePuy submission)
Cementless	£2,550.50		
Hybrid	£2,011.50		
Reverse Hybrid	£1,568.00		
All THR	£1,811.32	Weighted average of all THR	NA
RS	£1,029.00	Same as cemented THR	Cemented prostheses are the least costly, therefore lifetime costs are less likely to be over-estimated. Expert clinical opinion suggested approx. 90% of RS are performed with cement on the femoral side therefore costs are similar cemented THR
Surgical resource use (anaesthetic costs) (page 119 DePuy submission)			
Cemented	[redacted]	Based on data obtained from a leading NHS orthopaedic hospital in England and	Not provided
Cementless	[redacted]		
Hybrid	[redacted]		
Reverse Hybrid	[redacted]		

			validated by expert clinical opinion	
All THR	[redacted]		Weighted average based upon the number of hip primaries reported in the NJR 9th Annual Report	NA
RS	[redacted]		See cemented THR	See cemented THR
Surgical resource use (surgical consumables) (page 119 DePuy submission)				
Cemented	[redacted]		Based on data obtained from a leading NHS orthopaedic hospital in England and validated by expert clinical opinion	Not provided
Cementless	[redacted]			
Hybrid	[redacted]			
Reverse Hybrid	[redacted]			
All THR	[redacted]		Weighted average based upon the number of hip primaries reported in the NJR 9th Annual Report	NA
RS	[redacted]		See cemented THR	See cemented THR
Staff and theatre time (page 119 DePuy submission)				
	No. mins	Total cost		
Cemented	[redacted]	[redacted]	Micro-costing analysis	To provide an accurate assessment of the cost-differences
Cementless	[redacted]	[redacted]		
Hybrid	[redacted]	[redacted]		
Reverse Hybrid	[redacted]	[redacted]		

All THR	[redacted]	[redacted]		between different prosthesis classes as NHS Reference Costs do not disaggregate costs for procedures with cement and without cement. (page 115 DePuy submission)
RS	[redacted]	[redacted]		
†Assumed to equal hybrid. ‡Calculated based on the weighted average of THR procedures. §Assumed equal to cemented THR. Note figures have been rounded to 2 decimal places.				
Cost of primary procedure (page 121 DePuy submission)				
RS	[redacted]		See cemented THR	See cemented THR
Cemented THR	[redacted]		Micro-costing 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 without cement. (page 115 DePuy submission)
Cementless THR	[redacted]			
Hybrid THR	[redacted]			
Reverse Hybrid THR	[redacted]			
All THR	[redacted]			
DePuy [redacted]	[redacted]			
DePuy [redacted]	[redacted]			
Follow-up costs	£467.00		Department of Health (DoH) payment by results (PbR) 2012–13 tariff	Assumed to cover rehabilitation costs during the first three months post-surgery
Cost of revision	£13,399.42		Assumption that	Based on expert

		revision cost is double the mean cost of the primary procedure irrespective of class	opinion, the cost of revision surgery is considerably greater than the primary procedure

10.5.2 The base case results

THR versus RS in patients suitable for RS and THR

DePuy reported the following base case results for the comparison of THR with RS (page 124 DePuy submission) (Table 118 and Table 119).

Table 118. Base case results of DePuy submission considering THR vs. hip resurfacing in patients suitable for both procedures

Procedure	Costs	LYs	QALYs
All THR	£8,894	14.391	11.115
Resurfacing	£11,399	14.387	11.009
Difference	**-£2,504.31**	**0.004**	**0.106**
ICER	**All THR dominates**		

Table 119. Base case results for THR categories and RS of DePuy submission for patients suitable for THR and hip resurfacing

Technology	Costs	QALYs	[redacted]		vs. next less costly		Incremental analysis
			[redacted]	[redacted]	Inc. costs	Inc. QALYs	ICER
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cemented	£8,231	11.145	[redacted]	[redacted]	£20	-0.003	Dominated
Reverse hybrid	£8,570	11.148	[redacted]	[redacted]	£339	0.003	Dominated
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cementless	£8,743	11.146	[redacted]	[redacted]	£71	-0.007	Dominated
Hybrid	£8,817	11.167	[redacted]	[redacted]	£74	0.021	£33,338
All THR	£8,894	11.115	[redacted]	[redacted]	£77	-0.052	Dominated
Resurfacing	£11,399	11.009	[redacted]	[redacted]	£2,504	-0.106	Dominated

[redacted]

Abbreviations: ICER, incremental cost-effectiveness ratio; NSM, non-surgical management; QALY, quality-adjusted life years; THR, total hip replacement.

THR categories in patients not suitable for RS

The base case result for the comparison of the THR categories under investigation in patients not suitable for hip resurfacing arthroplasty showed that DePuy [redacted] was the most cost-effective intervention; dominating cemented THR, reverse hybrid THR, All THR, DePuy [redacted], cementless THR, and resurfacing (page 125 DePuy submission). Hybrid THR had an ICER of £780,788 compared with DePuy [redacted]. Table 120, taken from the manufacturer's submission, report the figures of this comparison.

Table 120. Base case results for THR categories of DePuy submission for patients not suitable for hip resurfacing arthroplasty

Technology	Costs	QALYs	[redacted]		vs. next less costly		Incremental analysis
			[redacted]	[redacted]	Inc. costs	Inc. QALYs	ICER
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cemented	£7,709	8.811	[redacted]	[redacted]	£13	-0.002	Dominated
Reverse hybrid	£8,158	8.805	[redacted]	[redacted]	£449	-0.006	Dominated
All THR	£8,198	8.801	[redacted]	[redacted]	£40	-0.004	Dominated
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cementless	£8,383	8.799	[redacted]	[redacted]	£117	-0.010	Dominated
Hybrid	£8,488	8.814	[redacted]	[redacted]	£104	0.015	£780,788
[redacted]							

Abbreviations: ICER, incremental cost-effectiveness ratio; NSM, non-surgical management; QALY, quality-adjusted life years; THR, total hip replacement

Results of the sensitivity analyses undertaken by DePuy

DePuy undertook five scenario analyses to *"…investigate the impact on the results of key methodological assumptions, including those relating to procedure costs, HRQoL, and the extrapolation of the NJR data set.*" (page 128 DePuy submission)

1) NHS reference costs for hip replacement procedures were used instead of costs from the micro study. The analysis identified hybrid THR as the optimal strategy at a willingness to pay (WTP) threshold of £20,000 per QALY for patient 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 HRQoL on the ICERs. In this analysis DePuy [redacted] was the optimal strategy at a WTP threshold of £20,000 per QALY.

3) An exponential model for the risk of revision was used in order to investigate the impact of transition probabilities which were independent of time. In this analysis the cost of hip resurfacing arthroplasty was substantially greater than any class of THR and DePuy [redacted] was the optimal strategy at a WTP threshold of £20,000 per QALY.

4) The impact of alternative Weibull models of revision stratified by age at primary procedure <70 years was investigated. DePuy [redacted] and [redacted] accrued the lowest costs and DePuy [redacted] 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 [redacted] was the most expensive class of THR and hybrid THR was the optimal strategy at a WTP threshold of £20,000 per QALY.

The results of the scenario analyses are reported in Table 121 and

Table 122.

Table 121. Results reported in the DePuy submission for the scenario analyses of patients suitable for hip resurfacing arthroplasty

	Base case		Scenario 1: NHS Reference Costs		Scenario 2: PROMS		Scenario 3: Exponential model		Scenario 4: Age <70 years model		Scenario 5: Age <55 years Model	
	Costs	QALYs	Costs	QALYs	Costs	QALYs	Costs	QALYs	Costs	QALYs	Costs	QALYs
[redacted]												
Cemented	£8,231	11.145	£7,642	11.145	£8,231	10.886	£8,476	11.126	£8,330	11.138	£8,454	11.127
Reverse hybrid	£8,570	11.148	£7,620	11.148	£8,570	10.889	£9,011	11.112	£8,595	11.147	£8,570	11.148
[redacted]												
Cementless	£8,743	11.146	£7,618	11.146	£8,743	10.886	£9,416	11.090	£8,831	11.138	£8,950	11.128
Hybrid	£8,817	11.167	£7,521	11.167	£8,817	10.907	£9,187	11.137	£8,872	11.163	£8,840	11.167
All THR	£8,894	11.115	£7,789	11.115	£8,894	10.855	£9,406	11.073	£9,048	11.102	£9,325	11.078
Resurfacing	£11,399	11.009	£10,087	11.009	£11,399	10.749	£11,560	10.997	£11,418	11.008	£11,569	10.999
[redacted]												
[redacted]												

[redacted]

Abbreviations: NHS, National Health Service; PROMs, patient reported outcome measures; QALY, quality-adjusted life year; THR total hip replacement

Table 122. Results reported in the DePuy submission for the scenario analyses of patients not suitable for hip resurfacing arthroplasty

Technology	Base case		Scenario 1: NHS Reference Costs		Scenario 2: PROMS		Scenario 3: Exponential model		Scenario 4: Age <70 years model		Scenario 5: Age <55 years model	
	Costs	QALYs	Costs	QALYs	Costs	QALYs	Costs	QALYs	Costs	QALYs	Costs	QALYs
[redacted]												
Cemented	£7,709	8.811	£7,321	8.811	£7,709	8.607	£7,823	8.804	£7,779	8.806	£8,209	8.772
Reverse hybrid	£8,158	8.805	£7,354	8.805	£8,158	8.600	£8,416	8.788	£8,323	8.792	£8,158	8.805
All THR	£8,198	8.801	£7,381	8.801	£8,198	8.596	£8,406	8.787	£8,292	8.794	£8,630	8.768
[redacted]												
Cementless	£8,383	8.799	£7,379	8.799	£8,383	8.595	£8,771	8.773	£8,373	8.801	£8,420	8.798
Hybrid	£8,488	8.814	£7,297	8.814	£8,488	8.609	£8,704	8.800	£8,528	8.811	£8,392	8.823
[redacted]												
[redacted]												
[redacted]												

Abbreviations: NHS, National Health Service; PROMs, patient reported outcome measures; QALY, quality-adjusted life year; THR total hip replacement

Univariate sensitivity analyses were carried out for patients suitable for THR and hip resurfacing arthroplasty producing a Tornado diagram for which key parameters were 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 form of 10,000 Monte Carlo simulations for patients who were suitable for both THR and hip resurfacing arthroplasty and for patients unsuitable for RS. For patients unsuitable for RS, PSA 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 resurfacing RS was associated with substantially higher costs and fewer QALYs compared with all classes of THR.

The manufacture's submission reported Table 123 and Table 124 with the results for the PSA.

Table 123. Results of the probabilistic sensitivity analysis by DePuy for patients suitable for THR and hip resurfacing arthroplasty

Technology	Costs	QALYs
[redacted]	[redacted]	[redacted]
Cemented	£8,240 (6,484 - 10,073)	11.145 (11.08 - 11.21)
Reverse-hybrid	£8,596 (6,740 - 10,450)	11.146 (11.07 - 11.22)
[redacted]	[redacted]	[redacted]
Cementless	£8,747 (7,068 - 10,482)	11.146 (11.08 - 11.21)
Hybrid	£8,826 (7,092 - 10,588)	11.166 (11.1 - 11.23)
Resurfacing	£11,408 (9,138 - 13,830)	11.009 (10.93 - 11.09)

[redacted]

Abbreviations: QALY, quality-adjusted d life years; THR, total hip replacement.

Table 124. Results of the probabilistic sensitivity analysis by DePuy for patients not suitable for hip resurfacing arthroplasty

Technology	**Costs**	**QALYs**
[redacted]	[redacted]	[redacted]
Cemented	£7,713 (6,118 - 9,409)	8.811 (8.76 - 8.86)
Reverse-hybrid	£8,171 (6,494 - 9,937)	8.804 (8.75 - 8.85)
[redacted]	[redacted]	[redacted]
Cementless	£8,387 (6,823 - 10,029)	8.799 (8.75 - 8.85)
Hybrid	£8,498 (6,872 - 10,216)	8.814 (8.76 - 8.86)

[redacted]

Abbreviations: QALY, quality-adjusted life years; THR, total hip replacement.

Strengths and weaknesses of the model

a) 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 (micro - costing analysis). By re-running the model, the review team could replicate the base case deterministic

and probabilistic results of the manufacturer's 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 with models in the literature which tend to use NHS reference costs which comprise all three cost components in a single value. NHS reference costs were subsequently investigated in sensitivity analyses.

b) Limitations

Limitations of the model identified concerned the following areas.

Revision rates:

Revision rates were modelled using a single Weibull fit that predicted a monotonic decreasing hazard through time. A bath tub hazard was briefly considered following Briggs et al. (2004).[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 four years after primary hip replacement and therefore it is likely a bath tub model could have been used. The submission acknowledges this is a limitation of the modelling. The manufacturer's probabilistic analyses 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. (2004)[38] and Higashi and Barendregt (2011).[272] This statement is misleading because the first two models used two Weibull fits (one to early and one to late failures) so as 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. (2003).[358] It should be noted that the figure for disutility was originally from a regression model output. Dawson et al. (2001)[294] reported the mean EQ-5D scores of 601 revision patients in the UK, following revision surgery the mean EQ-5D score at one year was

0.62. However, applying disutility (0.145) to the post-operative utility score does not reflect the lower QoL as reported in the original study (0.62 vs 0.635).

Resources and costs:

The cost-effectiveness of DePuy [redacted] and DePuy [redacted] were compared with the different THR and RS arthroplasty. In the base case analysis the costs were based on a micro-costing analysis and NHS reference costs were used in a scenario analysis, and it was assumed that all patients who received primary THR received a MoP articulation (regardless of whether they received a cemented, cementless or hybrid prosthesis). We agree with the manufacturer that the list price for DePuy products do not reflect the price available to the NHS, which results in uncertainty around their incremental cost/effectiveness ratio.

The variability of 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 costs, anaesthetic costs and theatre time reported in appendix E and H in the DePuy submission are all based on this micro-costing study undertaken by DePuy. Since the observational methods and the variance in resource use across the sample population was not reported in the submission, the review team was unable to verify the data. While undertaking a time and motion study to determine cost data inputs is desirable, to report a base case economic analysis on costs which cannot be verified is questionable.

DePuy assumed a unit cost of an in-patient stay of £295.29 basing the calculations on LoS data detailed in the NHS reference cost database (

Table 125). However, individual costs for the respective HRG codes were not reported and DePuy did not detail how they derived the costs for the weighted average length of stay which meant that the review team was unable to replicate the value used.

Table 125. Mean length-of-stay for patients receiving primary THR or hip resurfacing

HRG Name (Currency Code)	FCEs	LoS (days)	Unit cost per in-patient stay (£)/day	Source
Major Hip Procedures for non Trauma Category 1 with Major CC (HB12A)	2,573	9.92		NHS reference costs
Major Hip Procedures for non Trauma Category 1 with CC (HB12B)	6,433	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**	

HRG – healthcare resource group, FCEs- finished consultant episodes, LoS – length of stay

The review team clinical expert opinion suggests that the cost of revision surgery is greater than for primary THR/RS but revisions are carried out for a variety of reasons, and to assume that the cost of all revision procedures to be 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, where RS was not suitable, DePuy [redacted] was reported as the most cost effective intervention. However this result is dependent on allocation of relatively high cost of surgery with the Hybrid prosthesis based on micro-costing (Table 33 in the submission). However, no methodology was reported detailing how the model controlled for age and gender differences, while differences in both, age and gender, distributions were reported by DePuy (page 77 DePuy submission) (Table 126). 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 are desirable in order to compare THR and RS because of the dissimilarities among patient populations.

Table 126. Age and gender of patients receiving primary hip replacements in 2011†

	Cemented THR	Cementless THR	Other THR (e.g. hybrid)	Hip resurfacing	Total
Total procedures, n (%)	25,789 (36)	31,307 (44)	12,794 (18)	1,782 (2)	71,672
Total procedures with patient data, n (%)	4,739 (96)	29,751 (95)	12,241 (96)	1,600 (90)	68,331 (95)
Female, n (% of class)	16,112 (65)	16,731 (56)	7,743 (63)	241 (15)	40,827 (60)
Male, n (% of class)	8, 627 (35)	13,020 (44)	4,498 (37)	1,359 (85)	27,504 (40)
Average age, years	72.8	65.4	69.6	54.2	67.2
SD	9.7	11.3	10.9	9.5	13.4
Interquartile range	67.2 - 79.5	58.8 - 73.3	63.5 - 77.3	48.6 - 60.7	62.0 - 76.7

†Adapted from the Table 2.5 of the NJR 9th Annual report, 2012 (12)
THR - total hip replacement, SD - standard deviation

The base case probabilistic results are similar to those of 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.

In the base case analysis the manufacturer submission was largely in line with the NICE reference case. However, costs in the base-case analysis were not based on NHS reference costs, they were based on a micro-costing study. As mentioned earlier the micro-costing study could not be verified, however, the NHS reference cost estimates were based on a large sample size for both primary and revision surgery (n = 43,420 for primary and n = 26,797 for revision). Applying the NHS reference costs to both patient cohorts, the optimal strategy at a willingness to pay (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 data inputs which have been used.

The decision by the manufacture to not report cost – effectiveness acceptability curves (CEAC) in the main text is questionable. CEACs were included in the manufacturer's appendix and the reader 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 with costs from Vale et al. (2002),[19] as indicated in the manufacturer's submission, could not be identified in the report.

Given that the cost of revision only increased by 45% and not double, the cost of revision should have been tested using inflated Vale et al. (2002)[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 estimate 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 due to 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 gender differences, the lack of a cost-effectiveness analysis based on age and gender and the minimal reporting a CEAC. The main conclusion for the cost effectiveness evidence was that the DePuy devices are more cost effective compared to all other prosthesis. The hip RS arthroplasty was dominated by cemented THR, cementless THR, DePuy [redacted] DePuy [redacted], hybrid THR and reverse hybrid THR in patients suitable for both procedures. It was also noted that DePuy [redacted] 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 micro-costing analysis which was reported incompletely.

10.5.3 Smith & Nephew, Inc

Contents of submission

Smith & Nephew provided a 10 page non-systematic summary of the literature. They presented evidence 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 English and Australian National Joint Registers. They 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 clinical effectiveness review

a) Strengths:

The revision rates reported by bearing surface were extrapolated to 11 years.

b) 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. Resurfacing 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. The outcomes only consider revision following THR. Furthermore, the study and baseline population characteristics are not clearly presented and the results were 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 table of excluded studies. The cost-effectiveness review included a number of key papers but the list of included studies was not exhaustive probably due to the focus on revision.

Conclusions

The report is a subjective summary of the importance of bearing surfaces on revision rates and a justification to consider bearing surfaces in a sensitivity analysis within the cost-effectiveness model of the NICE report. It concludes that known factors to modify revision rates like bearing surfaces should be considered in analyses. It suggests that individual prostheses or design elements should be considered separately in analyses so that they impact on revision rates does

not get lost when grouping new technology implants for analysis. The information reported in this submission lacks objectivity and transparency.

10.5.4 Stryker

Contents of submission

Stryker provided a 22 page report which 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 resurfacing arthroplasty and summarised the complexity of available implants and corresponding revision rates considering evidence from the literature and National Joint Registries of England and Wales, Sweden, Norway and Australia. They concluded that the complexity of hip replacement procedures should be taken into consideration in economic evaluations and reported that Stryker is currently working with a group of researchers at the University of East Anglia and orthopedic 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 clinical effectiveness review

a) *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 included studies are not exhaustive. The review considered PROMs data for THR and revision rates were reported for three and eight years.

b) *Limitations:*

Stryker provided a non-systematic review of the clinical effectiveness literature using data referenced to the NJR 2011 annual report. The review lacks a clearly formulated research question 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. 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 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 identified considerable evidence on the cost-effectiveness of THR vs. THR.

Conclusions

Stryker did not answer a clearly formulated question but presented a summary of a selection of the available evidence. They provided details on the cemented procedure for the Exeter® stem (Stryker). They report 'very good' midterm results for the Exeter V40® stem. Stryker also reported results from the 2011 Australian Orthopedic Association National Joint Registry for the two stems listed above. Various published studies are listed which report positive results for these stems. The limited information reported in this submission lacks objectivity, transparency and clear conclusions.

10.5.5 JRI Orthopaedics Limited

Contents of submission

JRI 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 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 for all other cementless THRs, the majority of the hybrid and two of the six categories of the cemented THRs. Analysis of risk of revision by liner type and age showed that the risk of revision increased after the age of 70 when using a poly liner instead of a ceramic liner. Furthermore, a comparison of death rates of cemented versus cementless JRI implants

demonstrated a slightly higher death rate for patients receiving a cemented JRI THR in comparison to JRI cementless implants.

The JRI submission also included detailed clinical evaluation reports on five specific JRI brands detailing literature reviews and quality appraisals, four technical reports considering the JRI cemented and cementless components, coatings, details of the polyethylene used and specifications of the Trunion design. Finally, JRI submitted statistics of the NRJ 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 five years. The authors state that the majority of literature for their reviews was obtained online. Searches were undertaken in the Journal of Bone & Joint Surgery, Entrez Pubmed, the National Joint Register, and Google Scholar.

Limitations and strengths of clinical effectiveness review

a) Strengths:

The interventions, comparators and outcomes described by the manufacturer match those described in the final scope although the included studies are not exhaustive. The review includes 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 Furlong H-A.C THR, 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 between JRI cemented and cementless implants.

b) Limitations:

JRI provided only brief scoping reviews of the clinical effectiveness literature for each JRI brand but 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 a list of excluded studies. Study and baseline population characteristics are not clearly presented. The submission only briefly discusses revision rates of cemented THR compared to cementless THR from three national joint registries. The manufacturer's submission does not include a section on equity considerations.

Cost-effectiveness review – overall quality considerations

The submission provided very limited information concerned with the cost-effectiveness evidence.

Conclusions

JRI Orthopaedics Limited presented an overview of their 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.

11 Discussion

11.1 Decision problem and objectives

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

1. To compare the clinical and cost effectiveness of

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

11.2 Methods and summary of findings

Systematic Reviews

We undertook systematic reviews of 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 limited to studies published from 2008 and onward and sample sizes of 100 participants or more. Two independent reviewers screened all records, extracted data and independently assessed risk of bias Estimates of effectiveness were pooled and quality of evidence was assessed using the GRADE approach.

Although we appraised and summarized a very large amount of evidence much of it was inconclusive due to 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 versus 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 as opposed to a metal-on-polyethylene articulation, although the findings were inconclusive and reflected short-term follow-up.

Systematic reviews of cost-effectiveness and of 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 and cost-effectiveness and of registry data in order to identify inputs for the models to compare the clinical and cost-effectiveness of RS with different types of THR and different types of THR with each other.

National Joint Registry and Model Inputs

For the cost effectiveness analyses we used the national Joint Registry 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 inclusion of a further 5th mutually exclusive category. We identified time to revision for all categories by age and gender using NJR data and investigated a large number of methods for extrapolating beyond observed data and tested goodness of fit.
We built a Markov, multi state model to investigate both RS and THR. Health states included successful primary surgery, revision surgery, successful revision surgery and death after Fitzpatrick.[364] Cycle length was one year. We adopted a 10-year and a lifetime horizon analyses from the perspective of the NHS and personal and social services (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 ages and genders 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-gender matched pairs from the RS dataset and from all our five categories of THR combined. We used NHS Supply Chain for both RS and THR and sources from the literature for costs of follow-up and revision. We used age and gender adjusted utility values from the PROMs dataset using the same utility values for both procedures for before and after hip replacement and for revision, since no separate utility values were reported for RS.

We used age- and gender- 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 gender 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 over 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 where manufacturer costs were not available and alternative manufacturers' costs in sensitivity analyses.

We undertook sensitivity analyses and analyses of cost drivers including investigating changes in age and gender categories, stratifying by age (less than and more than 65 years) different methods of extrapolation of revision rates (using a lognormal model) and by varying prosthesis costs (using NHS list prices) and discount rates.

Findings

The National Joint Registry included just fewer than 420,000 patients. Just over 31,000 (7.4%) had undergone RS. Our identified categories of THR covered 62% of the THR population. Ninety percent of RS patients and 23% of THR category patients were less than 65 years old. Bathtub models (predicting increasing likelihood of revision over time) gave the best fit to the observed data. PROMs data showed that utility differences were dramatic – i.e., from pre-intervention at 0.35 to post intervention at 0.78 and from pre-revision at 0.53 to a similar level for post revision at 0.78. Revision rates varied; 97% of those undergoing THR, and 89% of those undergoing RS remained without revision by nine years.

RS vs. THR

Revision rates for all RS, compared to THR (all THR, all of our identified categories of THR combined, each of our THR categories separately) were always higher. The mean cost of RS was £2,672 and weighted mean cost of THR was £2,571.

Costs for RS were higher than for THR and mean QALYs gained were lower. The incremental cost-effectiveness ratio showed that RS was dominated by THR (For 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 deterministic and probabilistic results for RS

compared with THR and when analysed separately in sensitivity analyses for all age and gender groups. RS remained clearly dominated by THR. CEACs showed that for all patients, THR was almost 100% cost effective at any willingness to pay level.

THR vs. THR

The five categories of commonly used types of THR which we investigated are: Category A: CePoM (Cemented-cemented with a polyethylene-metal articulation) (125,285 patients); Category B: CeLPoM (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: HyPoM (Hybrid (cementless-cemented) with a polyethylene-metal articulation) (28,471 patients) and Category E: CePoC (Cemented-cemented with a polyethylene-ceramic articulation) (12,705 patients).

There were age and gender differences in the populations with different types of THR and variations in revision rates between Category E 1.6 % and Category C 3.5% at nine years. (For all interventions, revision rates at nine years were well under 10%). Prosthesis cost varied between £1,557.38 for Category A CePoM, and £3,868.80 for Category C CeLCoC.

For the base-case analysis, for all age and gender groups combined and using a bathtub model (indicating increasing likelihood of need for revision with time), and a lifetime horizon, Category E dominated the other four categories. Mean costs for Category E (CePoC) were slightly lower and mean QALYs for category E were slightly higher, than for all other THR categories in both deterministic and probabilistic analyses. In the deterministic analysis, compared to Category E, Category A (CePoM) cost £278 more (£14,801 compared to £14,523) and generated 0.0022 fewer QALYs (14.7887 as compared to 14.7909) and the probabilistic results were very similar. Over a lifetime horizon, Category E was 99.9% likely to be cost effective compared to Category A (1%) at a willingness to pay of £20,000 per QALY.

For patients aged over 65 years, at a willingness to pay of £20,000 per QALY, Category A was more likely to be cost effective in all groups aged over 65 years at a willingness to pay of £20,000 per QALY (Category A: 99% probability of being cost-effective; Categories B,C,D,E: <1% probability of being cost-effective), although Category E was more effective at a lifetime horizon for all groups (except for men aged 80 where the QALYs generated by Categories A and E were the same).

Sensitivity analyses using a lognormal model (indicating a decreasing risk of revision over time) gave category A as 99% cost-effective at a willingness to pay threshold of £20,000 per QALY and a lifetime horizon for all age-gender groups combined. Further sensitivity analysis using an age and gender adjusted lognormal model demonstrated at a lifetime horizon and a willingness to pay of £20,000 per QALY that likewise Category A was 100% cost-effective at a willingness to pay of £20,000 per QALY. The main drivers of differences between Categories A and E were found to be costs of components, discount rates and modelled revision rates.

11.3 Strengths and limitations

We undertook rigorous systematic reviews and we believe that we identified all relevant publications concerning the clinical and cost effectiveness of both THR and RS and available registry results. However, given the wide scope and large amount of identified evidence, we limited our inclusion for clinical effectiveness studies to a sample size of 100 and to studies published since 2008. This decision was based on our sample size calculations for clinically important differences in the Harris Hip Score and the fact that smaller studies tend to be underpowered to detect meaningful differences in continuous outcomes. We pooled data where possible and used the GRADE system for assessing overall quality.

We did not find any relevant longer term randomised controlled trials covering the comparison between RS and THR or between different types of THR which 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 was complete for patients' age and gender at receipt of THR.

However, the non-randomised nature of the database means that selection bias may be operating within the data. Revision rates may be higher e.g., for those undergoing RS because of an adverse profile in the population selected to receive one intervention rather than another). We worked to reduce confounding by propensity-matching RS with THR patients using NJR data and by undertaking extensive analyses by age and gender 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 frequency of use of different categories of prosthesis and in addition on expert clinical opinion. The selection of the five THR categories was conducted pre-hoc and prior to 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 hip replacement than previously generated and has the advantage of more precisely reflecting current practice.

We were only able to asses a relatively small number (5) of categories - as we needed to generate appropriate costings of sub components and to have large enough numbers 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 total hip replacements). However we modelled revision rates and survival using all hip replacements, for example to assess how our categories A-E compared in terms of revision rates and with RS. We found that overall revision rates were slightly higher overall. We found that overall revision rates were slightly higher than when we just investigated categories A-E. Given this finding we consider that our comparisons are likely to have focused on the more cost effective THR options.

The age and gender distributions varied between categories; when populations were controlled for differences in age and gender, or were stratified by gender and controlled for age, the lower revision rate for the CoPoC (E) category relative to other categories was not diminished. Also when well-fitting models which predicted either increasing or decreasing hazard on extrapolation were used, the superiority of the CoPoC (E) revision rate was again was upheld. There was insufficient information recorded consistently within the NJR for investigation of other potential confounders. For example, our clinical advisors suggested to us 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 which were reliably collected at the patient level in the NJR were age and gender. This means that we were unable to identify other characteristics or sub populations where RS might be more beneficial. However age and gender may act as a proxy for physicality and it is of interest that revision rates for RS were higher in every age and gender group 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 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. To model revision rates we followed NICE DSU recommendations in first exploring exponential, Weibull, Gompertz, lognormal and loglogistic models of observed revision rates based on IPD. However, previous economic analyses of hip replacement, notably those of Briggs et al. (2004)[38], Higashi et al. (2011)[272] and Pennington et al. (2013),[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 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.

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 over 85 years during the observation period revision rate was low and extrapolation with increasing hazard becomes less appropriate. We derived the bath tub hazard directly using the STATA package developed by Crowther and Lambert.[351] Pennington et al. (2013)[44] employed a piece-wise procedure to generate the "U" shaped hazard, however after extrapolation this predicted that more than 100% patients sustained revision and at this point the rate required capping. A strength of the work is that we tested a large number of methods for extrapolating revision including competing risks analysis and flexible parametric models.

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 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 differing THR head sizes. This demonstrated an advantage from a larger head size (36mm versus 28mm) and had a low risk of bias although so far follow up has only continued for one 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 gender differences. In our economic model we assumed both costs and utilities to be the same for the comparison of RS with THR. Our model is therefore likely to represent a fair comparison, but also is likely to underestimate the prosthesis cost of RS which has been reported as more expensive than THR prostheses. Edlin et al (2012).[40]. In spite of this assumption we found THR to be cost-effective (dominant) versus RS for all age (40, 50 and 60 years) and gender groups.

Though we undertook a rigorous systematic review for cost – effectiveness studies little information was available in the literature to estimate cost and resource usage. We could only identify one cost-utility analysis of RS versus THR from a randomised controlled trial. [40]. The costs of follow-up 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 life time of the model. This may have overestimated the cost of follow up although it was applied equally to both comparators in the model.

Costs of prosthesis varied between THR Categories. Category A was the least expensive but category E had lower revision rates and generated more QALYs over the life time horizon. We used prices for prosthesis components obtained from the NHS Supply Chain. We undertook a sensitivity analysis based on highest (Category A £1,557.38) and lowest (Category C, £3,868.80) prices. In order 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 costs of revision. We were unable to incorporate adverse events which 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 without the possibility for linking these. However, we were able to undertake sensitivity analyses to take account of possible costs 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 the structure in relation to these external comments.

12 CONCLUSIONS AND IMPLICATIONS FOR PRACTICE

THR 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. Overall, revision rates are low. However, although we appraised and summarized a very large amount of evidence much of the published literature was inconclusive due to poor reporting, missing data, inconsistent results and uncertainty in treatment effect estimates. Evidence on the relative benefits of RS versus THR or of different types of THR was largely lacking. Certain types of THR appeared to confer some benefit included 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.

12.1 RS vs. THR

Compared to THR, revision rates for RS were higher, mean costs for RS were higher and mean QALYs gained were lower; RS was dominated by THR.

Very similar results were obtained for deterministic and probabilistic results and for all age and gender groups and THR was almost 100% cost effective at any willingness to pay level.

12.2 THR vs. THR

Revision rates for all types of THR were low. Costs of prostheses varied depending partly on complexity (e.g. presence or absence of a liner, etc.) There were small but clear differences between categories in both costs and effectiveness as measured by QALYs and when age and gender groups were factored in. Category A was more cost effective for older age groups where revision rates are lower. However across all age gender groups combined, for the base-case analysis, mean costs for Category E (CePoC) were slightly lower and 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.

12.3 Recommendations for research

1. Randomised controlled trials with adequate length of 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 resources use data and implemented routinely would be extremely useful for future cost effectiveness assessments.
3. We would welcome work to validate our new findings on the relative cost-effectiveness of different combinations of prosthesis components for THR.

13 REFERENCES

1. Sinusas K. Osteoarthritis: Diagnosis and Treatment. *American Family Physician* 2012;85(1):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? *Clinical Science* 2012;123(11):627-34.

3. National Institute for Health and Clinical Excellence. *Osteoarthritis: The care and management of osteoarthritis in adults: NICE clinical guidelines CG59.* National Institute for Health and Care Excellence. 02/2008. URL:http://www.nice.org.uk/CG59 (accessed 18 April 2013).

4. National Institute for Health and Clinical Excellence. *Rheumatoid arthritis: the management of rheumatoid arthritis in adults: NICE clinical guidelines CG79.* National Institute for Health and Care Excellence. 02/2009. URL:http://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. *Annals of the Rheumatic Diseases* 2010;69(9):1580-8.

6. Aulakh TS, Kuiper JH, Dixey J, Richardson JB. Hip resurfacing for rheumatoid arthritis: independent assessment of 11-year results from an international register. *International Orthopaedics* 2011;35(6):803-8.

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. *Annals of the Rheumatic Diseases* 2010;69(4):631-7.

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.

9. Picavet HS, Hazes JM. Prevalence of self reported musculoskeletal diseases is high. *Annals of the Rheumatic Diseases* 2003;62(7):644-50.

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(2):e31521.

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 Technology Assessment* 2008;12(26):iii-223.

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(2):137-62.

13. Symmons D, Mathers C, Pfleger B. Global Burden of Osteoarthritis in the Year 2000. In: Global Burden of Diseases 2000 Geneva, Switzerland: World Health Organisation 2006.

14. Hoaglund FT, Steinbach LS. Primary osteoarthritis of the hip: etiology and epidemiology. *Journal of the American Academy of Orthopaedic Surgeons* 2001;9(5):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. *The Lancet* 2012;380(9855):1759-66.

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(9822):1199-204.

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(11):1270-85.

18. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. *Lancet* 2011;377(9783):2115-26.

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 Technology Assessment* 2002;6(15):1-109.

20. Buckwalter JA. Osteoarthritis and articular cartilage use, disuse, and abuse: experimental studies. *Journal of Rheumatology - Supplement* 1995;43:13-5.

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 & Rheumatism* 2000;43(3):568-75.

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. *Annals of Internal Medicine* 2000;133(8):635-46.

23. Kujala UM, Kaprio J, Sarna S. Osteoarthritis of weight bearing joints of lower limbs in former elite male athletes. *BMJ* 1994;308(6923):231-4.

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. *Journal of Bone & Joint Surgery - American Volume* 2004;86-A(7):1452-7.

25. National Institute for Health and Clinical Excellence. *Guidance on the use of metal on metal hip resurfacing arthroplasty: NICE technology appraisal guidance 44.* National Institute for Health and Care Excellence. 06/2002. URL:http://guidance.nice.org.uk/TA44/ (accessed 5 March 2013).

26. *OA Nation: the most comprehensive UK report of people with osteoarthritis.* Arthritis Care. 04/2004. URL:http://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-International Journal of the Care of the Injured* 2011;42(3):259-64.

28. Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S *et al*. WOMAC NRS 3.1 Osteoarthritis index delivered by mobile phone (m-WOMAC) is valid, reliable and responsive. *Internal Medicine Journal* 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 Musculoskeletal Disorders* 2003;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(2):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(4):737-55.

32. Merle D'Aubigné R, Postel M. Functional results of hip arthroplasty with acrylic prosthesis. *Journal of Bone & Joint Surgery - American Volume* 1954;36-A(3):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. *Scandinavian Journal of Rheumatology - Supplement* 1987;65:85-9.

34. Lequesne M. Indices of severity and disease activity for osteoarthritis. *Seminars in Arthritis & Rheumatism* 1991;20(6:Suppl 2):Suppl-54.

35. Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. *Journal of Rheumatology* 1997;24(4):779-81.

36. National Joint Registry. National Joint Registry for England and Wales: 8th Annual Report. Hemel Hempstead: The NJR Centre; 2011.

37. Jenkins PJ FAU, Clement ND FAU, Hamilton DF FAU, Gaston PF, Patton JT FAU, Howie CR. Predicting the cost-effectiveness of total hip and knee replacement: A health economic analysis. *Bone Joint J* 2013;95-B:115-21.

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(2):79-89.

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(3):315-26.

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(5):e001162.

41. Pivec R, Johnson K, Mears S, Mont MA. Hip Arthroplasty. *The Lancet* 2012;380(9855):1768-77.

42. *NHS atlas of variation in healthcare 2010.* South East Public Health Observatory. 29/11/2010. URL:http://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 Effectiveness & Resource Allocation* 2010;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.

45. Appleby J. *Which English hospital is best at hips?* The King's Fund. 29/09/2010. URL:http://www.kingsfund.org.uk/blog/2010/09/which-english-hospital-best-hips (accessed 22 November 2012).

46. National Institute for Health and Clinical Excellence. *Guidance on the selection of prostheses for primary total hip replacement: NICE technology appraisal*

guidance 2. National Institute for Health and Care Excellence. 04/2000. URL:http://www.nice.org.uk/TA2 (accessed 18 April 2013).

47. *Medical Device Alert: All metal-on-metal (MoM) hip replacements (MDA/2012/036).* Medicines and Health Products Regulation Agency. 25/06/2012. URL:http://www.mhra.gov.uk/Publications/Safetywarnings/MedicalDeviceAlerts/CON155761 (accessed 18 April 2013).

48. *National Joint Registry for England and Wales: 9th Annual Report.* National Joint Registry. 2012. URL:http://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. *Orthopaedic Nursing* 2012;31(5):296-9.

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 International* 2008;18(2):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. *Journal of Arthroplasty* 2003;18(7 Suppl 1):55-9.

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. *Journal of Bone & Joint Surgery - American Volume* 2005;87(8):1816-21.

53. Smith SL, Dowson D, Goldsmith AA. The effect of femoral head diameter upon lubrication and wear of metal-on-metal total hip replacements. *Proceedings of the Institution of Mechanical Engineers* 2001;215(2):161-70.

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. *Journal of Arthroplasty* 2004;19(8 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. *Journal of Bone & Joint Surgery - British Volume* 2011;93(8):1011-6.

56. *ASR Hip Recall Guide for Patients.* DePuy ASR Hip Recall Guide. 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. *Journal of Bone & Joint Surgery - American Volume* 2005;87(1):18-27.

58. Yan Y, Neville A, Dowson D. Understanding the role of corrosion in the degradation of metal-on-metal implants. *Proceedings of the Institution of Mechanical Engineers* 2006;220(2):173-81.

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. *Journal of Biomedical Materials Research* 1998;42(1):103-11.

60. Amstutz HC, Grigoris P. Metal on metal bearings in hip arthroplasty. *Clinical Orthopaedics & Related Research* 1996;329 Suppl:S11-S34.

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. *Journal of Biomedical Materials Research* 1996;30(2):157-63.

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. *Journal of Arthroplasty* 2004;19(8 Suppl 3):88-93.

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. *Journal of Bone & Joint Surgery - British Volume* 2008;90(7):847-51.

64. Case CP, Langkamer VG, James C, Palmer MR, Kemp AJ, Heap PF *et al*. Widespread dissemination of metal debris from implants. *Journal of Bone & Joint Surgery - British Volume* 1994;76(5):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. *Critical Reviews in Toxicology* 2008;38(8):645-74.

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(16):4477-83.

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. *Journal of Bone & Joint Surgery - British Volume* 2004;86(4):598-606.

68. Makela 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. *British Medical Journal* 2012;345: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.

70. 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. *Clinical Orthopaedics & Related Research* 2010;468(9):2357-62.

71. *National Medical Policy: Hip Resurfacing*. Health Net. 03/2013. URL:https://www.healthnet.com/static/general/unprotected/pdfs/national/policies/HipResurfacing.pdf (accessed 15 May 2013).

72. Johnson AJ, Zywiel MG, Hooper H, Mont MA. Narrowed indications improve outcomes for hip resurfacing arthroplasty. *Bulletin of the NYU Hospital for Joint Diseases* 2011;69 Suppl 1:S27-S29.

73. Havelin LI. The Norwegian Joint Registry. *Bulletin of the Hospital for Joint Diseases* 1999;58(3):139-47.

74. Bozic KJ, Kurtz SM, Lau E, Ong K, Vail TP, Berry DJ. The epidemiology of revision total hip arthroplasty in the United States. *Journal of Bone & Joint Surgery - American Volume* 2009;91(1):128-33.

75. Jafari SM, Coyle C, Mortazavi SM, Sharkey PF, Parvizi J. Revision hip arthroplasty: infection is the most common cause of failure. *Clinical Orthopaedics & Related Research* 2010;468(8):2046-51.

76. 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. *Orthopedic Clinics of North America* 2011;42(2):241-50.

77. 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. *Journal of Bone & Joint Surgery - British Volume* 2008;90(10):1291-7.

78. Parvizi J, Kim KI, Goldberg G, Mallo G, Hozack WJ. Recurrent instability after total hip arthroplasty: beware of subtle component malpositioning. *Clinical Orthopaedics & Related Research* 2006;447:60-5.

79. Parvizi J, Wade FA, Rapuri V, Springer BD, Berry DJ, Hozack WJ. Revision hip arthroplasty for late instability secondary to polyethylene wear. *Clinical Orthopaedics & Related Research* 2006;447:66-9.

80. Restrepo C, Mortazavi SM, Brothers J, Parvizi J, Rothman RH. Hip dislocation: are hip precautions necessary in anterior approaches? *Clinical Orthopaedics & Related Research* 2011;469(2):417-22.

81. 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. *Journal of Bone & Joint Surgery - British Volume* 2008;90(5):556-61.

82. Iannotti JP, Balderston RA, Booth RE, Rothman RH, Cohn JC, Pickens G. Aseptic loosening after total hip arthroplasty. Incidence, clinical significance, and etiology. *Journal of Arthroplasty* 1986;1(2):99-107.

83. 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. *Clinical Orthopaedics & Related Research* 2011;469(2):319-29.

84. Hartman CW, Garvin KL. Femoral fixation in revision total hip arthroplasty. *Journal of Bone & Joint Surgery - American Volume* 2011;93(24):2311-22.

85. Lindahl H, Garellick G, Regner H, Herberts P, Malchau H. Three hundred and twenty-one periprosthetic femoral fractures. *Journal of Bone & Joint Surgery - American Volume* 2006;88(6):1215-22.

86. 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 Orthopaedica Scandinavica* 2004;75(1):16-20.

87. Clift B. Periprosthetic fracture of the femur. *J Bone Joint Surg Am* 2000;82(3):446-7.

88. 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. *Journal of Arthroplasty* 2011;26(4):665-8.

89. 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. *Journal of Arthroplasty* 2010;25(8):1216-22.

90. Darwiche H, Barsoum WK, Klika A, Krebs VE, Molloy R. Retrospective analysis of infection rate after early reoperation in total hip arthroplasty. *Clinical Orthopaedics & Related Research* 2010;468(9):2392-6.

91. Bottner F, Su E, Nestor B, Azzis B, Sculco TP, Bostrom M. Radiostereometric Analysis: The Hip. *HSS Jrnl* 2005;1(1):94-9.

92. Lavernia CJ. Cost-effectiveness of early surgical intervention in silent osteolysis. *Journal of Arthroplasty* 1998;13(3):277-9.

93. *Australian Orthopaedic Association National Joint Replacement Registry: 2012 Annual Report.* Australian Orthopaedic Association National Joint Replacement Registry. 15/02/2013. URL:https://aoanjrr.dmac.adelaide.edu.au/annual-reports-2012 (accessed 18 April 2013).

94. Garellick G, Kärrholm J, Rogmark C, Herberts P. *Swedish Hip Arthroplasty Register: Annual Report 2010.* Swedish Hip Arthroplasty Register. 10/2011. URL:http://www.shpr.se/Libraries/Documents/AnnualReport-2010-2-eng.sflb.ashx (accessed 18 April 2013).

95. NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. CRD Report 4. York: NHS Centre for Reviews and Dissemination, University of York; 1999.

96. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.

97. 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.

98. 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(February):10.

99. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J *et al*. GRADE guidelines: 1. Introduction- GRADE evidence profoles and summary of findings. *J Clin Epidemiol* 2011;64(4):383-94.

100. 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. 2 nd Edition. CRD Report No. 4.* York: NHS Centre for Reviews and Dissemination (CRD), University of York; 2001.

101. DerSimonian R, Larid N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;7:177-88.

102. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* The Cochrane Collaboration. 03/2011. URL:http://handbook.cochrane.org/ (accessed 24 April 2013).

103. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(109):629-34.

104. 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 Musculoskeletal Disorders* 2010;11:8.

105. 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. *Annals of the Rheumatic Diseases* 2005;64(1):29-33.

106. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research* 2005;14(6):1523-32.

107. Bjorgul 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. *Journal of Orthopaedics & Traumatology* 2010;11(1):37-45.

108. Bjorgul 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. *Orthopaedics and Traumatology: Surgery and Research* 2010;96(2):97-103.

109. Angadi DS FAU, Brown S FAU - Crawfurd, 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(12):1604-10.

110. 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. *Journal of Arthroplasty* 2012;27(8 Suppl):2-7.

111. Engh Jr CA, Stepniewski AS, Ginn SD, Beykirch SE, Sychterz-Terefenko CJ, Hopper Jr RH *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(6 Suppl 2):17-25.

112. Capello WN, D'Antonio JA, Feinberg JR, Manley MT, Naughton M. Ceramic-on-ceramic total hip arthroplasty: update. *Journal of Arthroplasty* 2008;23(7:Suppl):39-43.

113. 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. *Clinical Orthopaedics & Related Research* 2005;(436):164-71.

114. D'Antonio J, Capello W, Manley M. Alumina ceramic bearings for total hip arthroplasty. *Orthopedics* 2003;26(1):39-46.

115. Mesko JW, D'Antonio JA, Capello WN, Bierbaum BE, Naughton M. Ceramic-on-Ceramic Hip Outcome at a 5- to 10-Year Interval. *Journal of Arthroplasty* 2011;26(2):172-7.

116. 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. *Journal of Bone & Joint Surgery - American Volume* 2011;93(14):1335-8.

117. 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. *Journal of Bone & Joint Surgery - American Volume* 2002;84-A(10):1823-8.

118. Bourne RB, Corten K. Cemented versus cementless stems: a verdict is in. *Orthopedics* 2010;33(9):638.

119. 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. *Clinical Orthopaedics & Related Research* 2011;469(1):209-17.

120. 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. *Journal of Bone & Joint Surgery - American Volume* 2012;94(12):1095-102.

121. 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(12 Suppl 2):37183.

122. 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. *Journal of Arthroplasty* 2011;26(6 Suppl):72-7.

123. 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. *Journal of Orthopaedic Research* 2011;29(8):1222-9.

124. Healy WL, Tilzey JF, Iorio R, Specht LM, Sharma S. Prospective, randomized comparison of cobalt-chrome and titanium trilock femoral stems. *Journal of Arthroplasty* 2009;24(6):831-6.

125. Kim YH, Choi Y, Kim JS. Comparison of bone mineral density changes around short, metaphyseal-fitting, and conventional cementless anatomical femoral components. *Journal of Arthroplasty* 2011;26(6):931-40.

126. 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. *Journal of Bone & Joint Surgery - British Volume* 2011;93(4):449-55.

127. 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.

128. 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. *Clinical Orthopaedics and Related Research* 2010;468(2):318-25.

129. 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 International* 2010;20(1):1-13.

130. Vendittoli PA, Lavigne M, Girard J, Roy AG. A randomised study comparing resection of acetabular bone at resurfacing and total hip replacement. *Journal of Bone & Joint Surgery - British Volume* 2006;88(8):997-1002.

131. Girard J, Lavigne M, Vendittoli PA, Roy AG. Biomechanical reconstruction of the hip: a randomised study comparing total hip resurfacing and total hip arthroplasty. *Journal of Bone & Joint Surgery - British Volume* 2006;88(6):721-6.

132. 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. *Journal of Arthroplasty* 2009;24(2):256-62.

133. 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 International* 2006;16 Suppl 4:73-81.

134. 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. *Journal of Arthroplasty* 2012;27(8):1544-53.

135. 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. *International Orthopaedics* 2011;35(8):1109-18.

136. Clement ND, Biant LC, Breusch SJ. Total hip arthroplasty: to cement or not to cement the acetabular socket? A critical review of the literature. *Archives of Orthopaedic and Trauma Surgery* 2012;132(3):411-27.

137. 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.

138. Yoshitomi H, Shikata S, Ito H, Nakayama T, Nakamura T. Manufacturers affect clinical results of THA with zirconia heads: A systematic review. *Clinical Orthopaedics and Related Research* 2009;467(9):2349-55.

139. 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. *Journal of Arthroplasty* 2011;26(3):419-26.

140. 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(6):684-95.

141. 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. *Journal of Arthroplasty* 2009;24(6):2-8.

142. Bhan S, Pankaj A, Malhotra R. One- or two-stage bilateral total hip arthroplasty: a prospective, randomised, controlled study in an Asian population. *Journal of Bone & Joint Surgery - British Volume* 2006;88(3):298-303.

143. 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? *Journal of Arthroplasty* 2009;24(1):71-6.

144. 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: Canadian Agency for Drugs and Technologies in Health (CADTH); 2008.

145. D'Arrigo C, Speranza A, Monaco E, Carcangiu A, Ferretti A. Learning curve in tissue sparing total hip replacement: comparison between different approaches. *Journal of Orthopaedics & Traumatology* 2009;10(1):47-54.

146. 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. *The Journal of bone and joint surgery American volume* 2007;89(6):1153-60.

147. Dutka J, Sosin P, Libura M, Skowronek P. Total hip arthroplasty through a minimally invasive lateral approach--our experience and early results. *Ortopedia Traumatologia Rehabilitacja* 2007;9(1):39-45.

148. Flivik G, Kristiansson I, Kesteris U, Ryd L. Is removal of subchondral bone plate advantageous in cemented cup fixation? A randomized RSA study. *Clinical Orthopaedics & Related Research* 2006;448:164-72.

149. Flivik G. Fixation of the cemented acetabular component in hip arthroplasty. *Acta Orthop* 2005;76(316 Suppl):3-30.

150. 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(3):372-8.

151. 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. *Clinical Orthopaedics & Related Research* 2011;469(1):200-8.

152. 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. *Journal of Bone & Joint Surgery - American Volume* 2003;85-A(8):1470-8.

153. Jolles BM, Bogoch ER. Posterior versus lateral surgical approach for total hip arthroplasty in adults with osteoarthritis. *Cochrane Database of Systematic Reviews* 2006;(3):CD003828.

154. Jolles BM, Bogoch ER. Surgical approach for total hip arthroplasty: direct lateral or posterior? *Journal of Rheumatology* 2004;31(9):1790-6.

155. 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. *Journal of Bone & Joint Surgery - British Volume* 2012;94(1):43-50.

156. Kim Y-H. Comparison of Primary Total Hip Arthroplasties Performed with a Minimally Invasive Technique or a Standard Technique. A Prospective and Randomized Study. *Journal of Arthroplasty* 2006;21(8):1092-8.

157. 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 International* 2007;17(1):4-8.

158. 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 International* 2006;16 Suppl 4:17-22.

159. Tang Z. Minimally invasive total hip replacement. Ottawa: Canadian Coordinating Office for Health Technology Assessment (CCOHTA); 2004. No. 4

160. Haverkamp D, Van Den Bekerom MPJ, Harmse I, Schafroth MU. One stage bilateral total hip arthroplasty, is it safe? A meta-analysis. *Hip International* 2010;20(4):440-6.

161. 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(5):538-44.

162. Li N, Deng Y, Chen L. Comparison of complications in single-incision minimally invasive THA and conventional THA. *Orthopedics* 2012;35(8):e1152-e1158.

163. 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. *Journal of Arthroplasty* 2011;26(8):1362-72.

164. 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. *Clinical Biomechanics* 2009;24(10):812-8.

165. Mazoochian F, Weber P, Schramm S, Utzschneider S, Fottner A, Jansson V. Minimally invasive total hip arthroplasty: a randomized controlled prospective trial. *Archives of Orthopaedic & Trauma Surgery* 2009;129(12):1633-9.

166. 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. *Archives of Orthopaedic & Trauma Surgery* 2012;132(5):725-31.

167. 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. *Archives of Orthopaedic & Trauma Surgery* 2011;131(2):179-89.

168. Nakamura N, Sugano N, Nishii T, Kakimoto A, Miki H. A comparison between robotic-assisted and manual implantation of cementless total hip arthroplasty. *Clinical Orthopaedics & Related Research* 2010;468(4):1072-81.

169. 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. *Journal of Bone & Joint Surgery - American Volume* 2005;87(4):701-10.

170. Pakvis D, Luites J, Hellemondt G, Spruit M. A cementless, elastic press-fit socket with and without screws. *Acta Orthop* 2012;83(5):481-7.

171. Parvizi J, Tarity TD, Sheikh E, Sharkey PF, Hozack WJ, Rothman RH. Bilateral total hip arthroplasty: One-stage versus two-stage procedures. *Clinical Orthopaedics & Related Research* 2006;453:137-41.

172. Pospischill M, Kranzl A, Attwenger B, Knahr K. Minimally invasive compared with traditional transgluteal approach for total hip arthroplasty: a comparative gait analysis. *Journal of Bone & Joint Surgery - American Volume* 2010;92(2):328-37.

173. Reininga IH, Wagenmakers R, van dA-S, 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 Musculoskeletal Disorders* 2007;8:4.

174. 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 Musculoskeletal Disorders* 2010;11:92.

175. Restrepo C, Parvizi J, Pour AE, Hozack WJ. Prospective randomized study of two surgical approaches for total hip arthroplasty. *Journal of Arthroplasty* 2010;25(5):671-9.

176. Saito S, Tokuhashi Y, Ishii T, Mori S, Hosaka K, Taniguchi S. One- versus two-stage bilateral total hip arthroplasty. *Orthopedics* 2010;33(8).

177. 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. *Orthopaedics & traumatology, surgery & research* 2012;98(2):151-8.

178. 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. *Journal of Arthroplasty* 2012;27(5):687-94.

179. 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. *European Journal of Medical Research* 2009;14(6):256-63.

180. 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(5):e37947.

181. 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. *Clinical Orthopaedics & Related Research* 2011;469(4):1110-8.

182. 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. *Clinical Orthopaedics & Related Research* 2010;468(2):565-70.

183. Brodner W, Bitzan P, Meisinger V, Kaider A, Gottsauner-Wolf F, Kotz R. Serum cobalt levels after metal-on-metal total hip arthroplasty. *Journal of Bone & Joint Surgery - American Volume* 2003;85-A(11):2168-73.

184. 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(4):549-58.

185. 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(4):559-66.

186. 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(2):218-26.

187. 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. *Journal of Orthopaedic Research* 2004;22(5):1035-43.

188. 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. *Journal of Bone & Joint Surgery - American Volume* 2004;86-A(5):963-74.

189. 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(5):719-25.

190. 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. *Journal of Orthopaedic Research* 2005;23(6):1280-5.

191. 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. *Journal of Bone & Joint Surgery - British Volume* 2006;88(8):1003-5.

192. Hartl A, Schillinger M, Wanivenhaus A. Cemented versus cementless total hip arthroplasty for osteoarthrosis and other non-traumatic diseases. *Cochrane Database of Systematic Reviews* 2004;(3):CD004850.

193. Howie DW, McGee MA, Costi K, Graves SE. Metal-on-metal resurfacing versus total hip replacement-the value of a randomized clinical trial. *Orthopedic Clinics of North America* 2005;36(2):195-201.

194. 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. *Journal of Arthroplasty* 2004;19(4):469-75.

195. 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. *Journal of Bone & Joint Surgery - American Volume* 2002;84-A(9):1651-8.

196. 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. *Journal of Bone and Joint Surgery - Series B* 2007;89(2):174-9.

197. 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(8):1769-76.

198. Kraay MJ, Thomas RD, Rimnac CM, Fitzgerald SJ, Goldberg VM. Zirconia versus Co-Cr femoral heads in total hip arthroplasty: early assessment of wear. *Clinical orthopaedics and related research* 2006;453:86-90.

199. Lombardi J, Mallory TH, Cuckler JM, Williams J, Berend KR, Smith TM. Mid-term results of a polyethylene-free metal-on-metal articulation. *Journal of Arthroplasty* 2004;19(7 SUPPL 2):42-7.

200. 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. *Clinical Orthopaedics and Related Research* 2003;(406):282-96.

201. 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(3):344-9.

202. Medical Advisory Secretariat. Metal-on-Metal Total Hip Resurfacing Arthroplasty: An Evidence-Based Analysis. *Ontario Health Technology Assessment Series* 2006;6(4):1-57.

203. Meek RD, Allan DB. Cemented versus cementless surgical approach for total hip arthroplasty revision. *Cochrane Database of Systematic Reviews* 2005;(2):CD005322.

204. Alberta Heritage Foundation for Medical Research. Metal-on-metal hip resurfacing for young, active adults with degenerative hip disease. Edmonton: Alberta Heritage Foundation for Medical Research (AHFMR); 2002. No. Technote TN 33

205. 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(4):479-84.

206. Pitto RP, Blanquaert D, Hohmann D. Alternative bearing surfaces in total hip arthroplasty: zirconia-alumina pairing. Contribution or caveat? *Acta Orthopaedica Belgica* 2002;68(3):242-50.

207. Pitto RP, Schikora N, Willmann G, Graef B, Schmidt R. Radiostereoanalysis of press-fit cups with alumina liner. - A randomized clinical trial. *Bioceramics 15* 2003;240-2:817-21.

208. 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 International* 2005;15(3):129-35.

209. 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(1):71-8.

210. 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. *Journal of Bone & Joint Surgery - British Volume* 2005;87(12):1631-5.

211. 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 Services Research* 2004;4(1):39.

212. 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 and Cartilage* 2007;15(9):981-1000.

213. Zhou ZK, Li MG, Borlin N, Wood DJ, Nivbrant B. No increased migration in cups with ceramic-on-ceramic bearing: an RSA study. *Clinical Orthopaedics & Related Research* 2006;448:39-45.

214. 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. *Journal of Arthroplasty* 2009;24(6 Suppl):9-14.

215. Cai P, Hu Y, Xie J. Large-diameter Delta Ceramic-on-ceramic Versus Common-sized Ceramic-on-polyethylene Bearings in THA. *Orthopedics* 2012;35(9):e1307-e1313.

216. 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. *Journal of Arthroplasty* 2009;24(6):837-45.

217. 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(6):746-54.

218. 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.

219. 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. *Clinical Orthopaedics & Related Research* 2004;(429):6-16.

220. 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. *Clinical Orthopaedics & Related Research* 2012;470(11):3083-93.

221. 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. *Journal of Arthroplasty* 2008;23(3):337-43.

222. 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. *Journal of Bone & Joint Surgery - American Volume* 2011;93(8):716-22.

223. 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(1):6-12.

224. 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(9):1108-16.

225. 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. *Clinical Orthopaedics & Related Research* 2010;468(2):326-36.

226. 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. *Journal of Arthroplasty* 2010;25(3):392-7.

227. 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. *Journal of Bone & Joint Surgery - British Volume* 2011;93(12):1602-9.

228. 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. *Journal of Bone and Joint Surgery-British Volume* 2010;92B(7):935-40.

229. 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. *Journal of Bone & Joint Surgery - British Volume* 2012;94(4):459-65.

230. 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(4):410-6.

231. 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. *International Orthopaedics* 2011;35(5):667-74.

232. 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. *Journal of Bone & Joint Surgery - British Volume* 2012;94(11):1462-7.

233. 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(5):559-66.

234. Smolders JM FAU, 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(11):1509-14.

235. 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 Chirurgiae Orthopaedicae et Traumatologiae Cechoslovaca* 2011;78(5):410-5.

236. 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? *International Orthopaedics* 2011;35(12):1771-6.

237. 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 Musculoskeletal Disorders* 2008;9:136.

238. 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. *Journal of Bone & Joint Surgery - American Volume* 2009;91(4):773-82.

239. Voleti PB, Baldwin KD, Lee GC. Metal-on-Metal vs Conventional Total Hip Arthroplasty: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Journal of Arthroplasty* 2012;27(10):1844-9.

240. Pailhe R, Sharma A, Reina N, Cavaignac E, Chiron P, Laffosse JM. Hip resurfacing: a systematic review of literature. *International Orthopaedics* 2012;36(12):2399-410.

241. 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. *International Journal Of Surgery* 2013;11(6):477-82.

242. 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. *Orthopedic Reviews* 2013;5(1):e8.

243. 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.

244. Dechartres A, Trinquart L, Boutron I, avaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ* 2013;346:f2304.

245. Kalairajah Y, Azurza K, Hulme C, Molloy S, Drabu KJ. Health outcome measures in the evaluation of total hip arthroplasties--a comparison between the Harris hip score and the Oxford hip score. *Journal of Arthroplasty* 2005;20(8):1037-41.

246. Glanville J, Kaunelis D, Mensinkai S. How well do search filters perform in identifying economic evaluations in MEDLINE and EMBASE. *International Journal of Technology Assessment in Health Care* 2009;25(4):522-9.

247. Paisley S, Booth A, Mensinkai S. *Health-related quality of life studies. In: Etext on Health Technology Assessment (HTA) Information Resources.* National Library of Medicine. 2005. URL:http://www.nlm.nih.gov/archive/20060905/nichsr/ehta/chapter12.html (accessed 6 March 2013).

248. 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. *International Journal of Technology Assessment in Health Care* 2005;21(2):240-54.

249. 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(4):355-71.

250. 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. *Journal of Bone & Joint Surgery, British Volume* 2008;90-B(SUPP III):504.

251. 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. *Journal of Bone and Joint Surgery-British Volume* 2011;93B(2):158-63.

252. 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? *Clinical orthopaedics and related research* 2010;468(9):2301-12.

253. Amman S, Cizik A, Leopold SS, Manner PA. Two-incision minimally invasive vs standard total hip arthroplasty: comparison of component position and hospital costs. *Journal of Arthroplasty* 2012;27(8):1569-74.

254. Duwelius PJ, Brenner JS, Reyner DP, George JC. Cost Effectiveness of Minimally Invasive Total Hip Arthroplasty. *Seminars in Arthroplasty* 2008;19(2):186-93.

255. 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 International* 2006;16 Suppl 4:54-7.

256. 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(3):e000752.

257. 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(5):609-17.

258. 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. *Journal of Bone and Joint Surgery* 2006;88(4):706-14.

259. Marinelli M, Soccetti A, Panfoli N, de PL. Cost-effectiveness of cemented versus cementless total hip arthroplasty. A Markov decision analysis based on implant cost. *Journal of Orthopaedics & Traumatology* 2008;9(1):23-8.

260. 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. *Journal of Surgical Research* 2011;169(2):227-33.

261. 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. *Journal of Bone & Joint Surgery - American Volume* 2009;91(3):634-41.

262. 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(2):149-59.

263. 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. *Internal Medicine Journal* 2012;42(6):620-6.

264. 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. *Medical Care* 2005;43(10):1018-26.

265. Bak P, Muller WD, Bocker B, Smolenski UC. Short-term patterns of recovery from total hip and knee arthroplasty after multidisciplinary inpatient rehabilitation. *Physikalische Medizin Rehabilitationsmedizin Kurortmedizin* 2008;18(1):11-8.

266. 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 Nutrition* 2009;12(8):1122-32.

267. Jibodh SR, Gurkan I, Wenz JF. In-hospital outcome and resource use in hip arthroplasty: Influence of body mass. *Orthopedics* 2004;27(6):594-601.

268. Segal L, Day SE, Chapman AB, Osborne RH. Can we reduce disease burden from osteoarthritis? *Medical Journal of Australia* 2004;180(5 Suppl):S11-S17.

269. 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. *International Orthopaedics* 2009;33(4):949-54.

270. Dutka J, Dutka L, Janiszewski M, Hajduk G. Cost analysis and sociomedical aspects of the conservative and surgical treatment of hip osteoarthritis. *Ortopedia Traumatologia Rehabilitacja* 2008;10(6):537-46.

271. 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(7):848-55.

272. 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(9):e25403.

273. 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 Musculoskeletal Disorders* 2011;12:43.

274. Mota REM. Cost-Effectiveness Analysis of Early versus Late Total Hip Replacement in Italy. *Value in Health* 2013;16(2):267-79.

275. Scheerlinck T, Duquet W, Casteleyn P-P. Socioeconomic aspects of total hip arthroplasty: A one-year survey in a Belgian university hospital. *Acta Orthopaedica Belgica* 2004;70(6):525-33.

276. Rana AJ, Iorio R, Healy WL. Hospital economics of primary THA decreasing reimbursement and increasing cost, 1990 to 2008. *Clinical Orthopaedics & Related Research* 2011;469(2):355-61.

277. O'Shea K, Bale E, Murray P. Cost analysis of primary total hip replacement. *Irish Medical Journal* 2002;95(6):177-80.

278. Zhang Y, Zhang H, Clarke HD, Hattrup SJ. Analysis of total joint arthroplasty costs in Chinese patients. *Journal of Arthroplasty* 2012;27(8):1423-8.

279. Stargardt T. Health service costs in Europe: Cost and reimbursement of primary hip replacement in nine countries. *Health Economics* 2008;17(1 Suppl):S9-S20.

280. 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(6):659-67.

281. Fielden JM, Cumming JM, Horne JG, Devane PA, Slack A, Gallagher LM. Waiting for hip arthroplasty: economic costs and health outcomes. *Journal of Arthroplasty* 2005;20(8):990-7.

282. 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(9):1144-50.

283. Bozic KJ, Chiu VW, Slover JD, Immerman I, Kahn JG. Health state utility in patients with osteoarthritis of the hip and total hip arthroplasty. *Journal of Arthroplasty* 2011;26(6 Suppl):129-32.

284. 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. *Journal of Arthroplasty* 2012;27(2):232-7.

285. March L, Cross M, Tribe K, Lapsley H, Courtenay B, Brooks P. Cost of joint replacement surgery for osteoarthritis: the patients' perspective. *Journal of Rheumatology* 2002;29(5):1006-14.

286. 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. *Scandinavian Journal of Caring Sciences* 2009;23(1):11-20.

287. de Palma L, Procaccini R, Soccetti A, Marinelli M. Hospital cost of treating early dislocation following hip arthroplasty. *Hip International* 2012;22(1):62-7.

288. Burns AWR, Bourne RB. Economics of revision total hip arthroplasty. *Current Orthopaedics* 2006;20(3):203-7.

289. 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 Journal of Surgery* 2008;78(10):875-80.

290. Kurtz SM, Lau E, Watson H, Schmier JK, Parvizi J. Economic burden of periprosthetic joint infection in the United States. *Journal of Arthroplasty* 2012;27(8 Suppl):61-5.

291. 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. *Journal of Bone & Joint Surgery, British Volume* 2012;94-B(5):619-23.

292. Parvizi J, Pawasarat IM, Azzam KA, Joshi A, Hansen EN, Bozic KJ. Periprosthetic joint infection: The economic impact of methicillin-resistant infections. *Journal of Arthroplasty* 2010;25(6 Suppl):103-7.

293. 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. *Quality of Life Research* 2004;13(10):1659-70.

294. 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. *Journal of Bone & Joint Surgery, British Volume* 2001;83-B(8):1125-9.

295. Jones CA, Pohar S. Health-related quality of life after total joint arthroplasty: a scoping review. *Clinics in Geriatric Medicine* 2012;28(3):395-429.

296. 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. *Journal of Bone & Joint Surgery - British Volume* 2004;86(6):801-8.

297. Rolfson O, Karrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: Results of a nationwide prospective observational study. *Journal of Bone & Joint Surgery, British Volume* 2011;93-B(7):867-75.

298. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. New York: Oxford University Press; 2005.

299. McKenzie L, Vale L, Stearns S, McCormack K. Metal on metal hip resurfacing arthroplasty. An economic analysis. *European Journal of Health Economics* 2003;4(2):122-9.

300. Akiyama H, Hoshino A, Iida H, Shindo H, Takakura Y, Miura H *et al*. A pilot project for the Japan arthroplasty Register. *Journal of Orthopaedic Science* 2012;17(4):358-69.

301. Buergi ML, Walter WL. Hip resurfacing arthroplasty: the Australian experience. *Journal of Arthroplasty* 2007;22(7 Suppl 3):61-5.

302. Conroy JL, Whitehouse SL, Graves SE, Pratt NL, Ryan P, Crawford RW. Risk factors for revision for early dislocation in total hip arthroplasty. *Journal of Arthroplasty* 2008;23(6):867-72.

303. Corten K, MacDonald SJ. Hip resurfacing data from national joint registries: what do they tell us? What do they not tell us? *Clinical Orthopaedics & Related Research* 2010;468(2):351-7.

304. Luo R, Brekke A, Noble PC. The financial impact of joint registries in identifying poorly performing implants. *Journal of Arthroplasty* 2012;27(8 Suppl):66-71.

305. Sexton SA, Walter WL, Jackson MP, De SR, Stanford T. Ceramic-on-ceramic bearing surface and risk of revision due to dislocation after primary total hip replacement. *Journal of Bone & Joint Surgery - British Volume* 2009;91(11):1448-53.

306. 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. *Journal of Bone and Joint Surgery - Series B* 2006;88(10):1293-8.

307. 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. *Journal of Bone & Joint Surgery - British Volume* 2012;94(12):1611-7.

308. 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.

309. Meek RMD, Allan DB, McPhillips G, Kerr L, Howie CR. Late dislocation after total hip arthroplasty. *Clinical Medicine and Research* 2008;6(1):17-23.

310. Stea S, Bordini B, De CM, Petropulacos K, Toni A. First hip arthroplasty register in Italy: 55,000 cases and 7 year follow-up. *International Orthopaedics* 2009;33(2):339-46.

311. 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(1):57-70.

312. 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(6):853-65.

313. 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(1):28-41.

314. 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. *Journal of Bone & Joint Surgery - British Volume* 2008;90(12):1562-9.

315. 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. *Journal of Bone and Joint Surgery-American Volume* 2008;90A(10):2160-70.

316. 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(1):42-52.

317. 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. *Journal of Bone & Joint Surgery - American Volume* 2011;93(2):178-86.

318. 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(5):521-9.

319. 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(3):207-13.

320. 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):1-5.

321. 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(4):1-11.

322. Espehaug B, Furnes O, Engesaeter LB, Havelin LI. Hip arthroplasty in Norway 1989-2008. *Tidsskrift for Den Norske Laegeforening* 2011;131(16):1543-8.

323. 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(6):649-59.

324. 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. *Clinical Orthopaedics & Related Research* 2012;470(11):3007-13.

325. 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 & research* 2010;62(4):473-9.

326. Franklin PD, Allison JJ, Ayers DC. Beyond joint implant registries: A patient-centered research consortium for comparative effectiveness in total joint replacement. *JAMA - Journal of the American Medical Association* 2012;308(12):1217-8.

327. Namba RS, Inacio MC, Paxton EW. Risk factors associated with surgical site infection in 30,491 primary total hip replacements. *Journal of Bone & Joint Surgery - British Volume* 2012;94(10):1330-8.

328. Paxton EW, Inacio M, Slipchenko T, Fithian DC. The kaiser permanente national total joint replacement registry. *Permanente Journal* 2008;12(3):12-6.

329. 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. *Journal of Bone & Joint Surgery - British Volume* 2006;88(10):1303-8.

330. 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(2):182-9.

331. 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. *Journal of Bone & Joint Surgery - British Volume* 2011;93(2):172-7.

332. 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(1):60-5.

333. 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(1):53-9.

334. 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(3):214-9.

335. Necas L, Katina S, Krivanek S, Uhlarova J. Slovakian Arthroplasty Register. Review of the annual report of the Slovakian Arthroplasty Register - 2010. *Acta Chirurgiae Orthopaedicae et Traumatologiae Cechoslovaca* 2011;78(Suppl B):1-59.

336. Boyer P, Boutron I, Ravaud P. Scientific production and impact of national registers: the example of orthopaedic national registers. *Osteoarthritis & Cartilage* 2011;19(7):858-63.

337. Graves SE, Rothwell A, Tucker K, Jacobs JJ, Sedrakyan A. A multinational assessment of metal-on-metal bearings in hip replacement. *Journal of Bone & Joint Surgery - American Volume* 2011;93 Suppl 3:43-7.

338. 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(4):393-401.

339. 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(5):535-41.

340. 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. *International Orthopaedics* 2012;36(11):2197-203.

341. 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. *International Orthopaedics* 2012;36(7):1349-54.

342. 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). *American Journal of Orthopedics (Chatham, Nj)* 2010;39(9 Suppl):5-13.

343. 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. *Journal of Bone & Joint Surgery - British Volume* 2012;94(6):746-54.

344. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70(1):41-55.

345. Leuven, E. and Sianesi, B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. (S432001). 2003. Boston College Department of Economics. Statistical Software Components.

346. StataCorp. Stata Statistical Software: Release 12.0. 2011. Texas: Stata Corporation.

347. Coviello V, Boggess M. Cumulative incidence estimation in the presence of competing risks. *The Stata Journal* 2004;4:103-12.

348. Collet D. *Modelling survival data in medical research*. Chapman and Hall/CRC; 2003.

349. *Design and analysis of clinical trials with time to event endpoints*. Chapman and Hall/CRC; 2009.

350. Machin D, Cheung YB, Parmar MKB. *Survival analysis: a practical approach*. Chichester: John Wiley and Sons Ltd; 2006.

351. Crowther, M. J. and Lanbert, P. C. stgenreg: A Stata package for general parametric survival analysis. *Journal of Statistical Software.* In Press 2013.

352. 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. *Statist Med* 2002;21(15):2175-97.

353. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata Journal* 2009;9(2):265-90.

354. Boyer KA, Beaupre GS, Andriacchi TP. Gender differences exist in the hip joint moments of healthy older walkers. *J Biomech* 2008;41(16):3360-5.

355. *Consideration of consultation responses on review proposal: Review of TA2; Hip disease replacement prostheses and TA44; Hip disease metal on metal resurfacing.* National Institute for Health and Care Excellence. 27/07/2011. URL:http://www.nice.org.uk/nicemedia/live/11386/55850/55850.pdf (accessed 19 June 2013).

356. Latimer NR. Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient level data: inconsistencies, limitations and a practical guide. *Medical Decision Making* 2013;0272989X12472398.

357. 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. 03/2013. URL:http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.pdf (accessed 9 April 2013).

358. 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.* University of York. 12/2003. URL:http://www.york.ac.uk/media/che/documents/papers/technicalpapers/CHE%20Technical%20Paper%2028.pdf (accessed 16 May 2013).

359. 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. *Journal of Bone & Joint Surgery - American Volume* 1993;75(7):961-75.

360. 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. *Journal of Bone & Joint Surgery - American Volume* 1997;79(1):53-64.

361. Callaghan JJ, Albright JC, Goetz DD, Olejniczak JP, Johnston RC. Charnley total hip arthroplasty with cement. Minimum twenty-five-year follow-up. *Journal of Bone & Joint Surgery - American Volume* 2000;82(4):487-97.

362. 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. *Journal of Bone & Joint Surgery - American Volume* 2004;86-A(4):690-5.

363. 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. *Journal of Bone & Joint Surgery - British Volume* 2004;86(4):492-7.

364. Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M. Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses. *Health Technology Assessment* 1998;2(20):1-64.

365. National Institute for Health and Clinical Excellence. *Guide to the Methods of Technology Appraisal.*2008. URL:http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed 29 May 2012).

366. *New Zealand Joint Registry: Thirteen Year Report January 1999 to December 2011.* HEALTHFIRST: Canterbury District Health Board's gateway to health information for everyone. 2012. URL:http://www.cdhb.govt.nz/njr/reports/A2D65CA3.pdf (accessed 16 May 2013).

367. Curtis L. Unit Costs of Health and Social Care 2011. University of Kent: Personal Social Services Research Unit; 2011.

368. *Patient Reported Outcome Measures (PROMS).* Health and Social Care Information Centre. 2013. URL:http://www.hscic.gov.uk/proms (accessed 16 May 2013).

369. 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(2):155-60.

370. Baad-Hansen T, Kold S, Olsen N, Christensen F, Soballe K. Excessive distal migration of fiber-mesh coated femoral stems. *Acta Orthop* 2011;82(3):308-14.

371. Bernath V. Hip resurfacing in patients with osteoarthritis. Clayton, Victoria: Centre for Clinical Effectiveness (CCE); 2002. No. 9

372. Beswick A, Wylde V, Blom A, Gooberman-Hill R, Dieppe PA. Pain After Hip or Knee Joint Replacement for Osteoarthritis: A Systematic Review. *Arthritis and Rheumatism* 2011;63(10):S413.

373. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* 2012;2(1):e000435.

374. 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(2):161-5.

375. Boden H, Adolphson P. No adverse effects of early weight bearing after uncemented total hip arthroplasty: a randomized study of 20 patients. *Acta Orthopaedica Scandinavica* 2004;75(1):21-9.

376. 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(1):13-9.

377. Butler R.A., Rosenzweig S, Myers L, Barrack RL. The Frank Stinchfield Award: the impact of socioeconomic factors on outcome after THA: a prospective, randomized study. *Clinical Orthopaedics & Related Research* 2011;469(2):339-47.

378. Cobb J. The Functional Outcome of Hip Resurfacing and Large-head THA is the Same A Randomized, Double-blind Study. *Clinical Orthopaedics and Related Research* 2010;468(11):3134.

379. 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(10):e13520.

380. D'Angelo F, Murena L, Zatti G, Cherubino P. The unstable total hip replacement. *Indian Journal of Orthopaedics* 2008;42(3):252-9.

381. Digas G, Karrholm J, Thanner J. Addition of fluoride to acrylic bone cement does not improve fixation of a total hip arthroplasty stem. *Clinical Orthopaedics and Related Research* 2006;448:58-66.

382. 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. *Journal of Arthroplasty* 2005;20(6):784-92.

383. 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 in Health* 2011;14(7):A245.

384. Eingartner C, Piel S, Weise K. Results of a cemented straight titanium alloy femoral stem after mean follow-up of 13 years. *European Journal of Orthopaedic Surgery and Traumatology* 2007;17(6):587-93.

385. 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 and Rheumatism* 2008;58(9):S245-S246.

386. Fick DP, Nivbrant B. Minimally invasive surgical approaches for total hip arthroplasty in adults with osteoarthritis. *Cochrane Database of Systematic Reviews* 2004;(2):CD004798.

387. 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 International* 2003;13(3):142-7.

388. 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 International* 2005;15(3):143-8.

389. 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. *Journal of Bone and Joint Surgery Series B* 2006;88(9):1143-8.

390. Hamadouche M, Baque F, Lefevre N, Kerboull M. Minimum 10-year survival of Kerboull cemented stems according to surface finish. *Clinical Orthopaedics & Related Research* 2008;466(2):332-9.

391. 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(4):417-22.

392. HAYES Inc. Ganz trochanteric flip osteotomy approach to hip resurfacing for treatment of osteoarthritis. Lansdale, PA: HAYES, Inc; 2012.

393. 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 of Systematic Reviews* 2008;(4):CD007494.

394. 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. *Journal of Orthopaedics & Traumatology* 2010;11(1):29-35.

395. 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. *Journal of Arthroplasty* 2009;24(8):1216-20.

396. ISPOR 14th Annual European Congress Research Abstracts. *Value in Health* 2011;14(7):A233-A574.

397. 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. *Technology & Health Care* 2008;16(3):195-214.

398. Jandric S, Jovicic Z, Novakovic S. Differences in pain between women and men in patients with total hip arthroplasty. *European Journal of Pain* 2009;13 Suppl 1:S133.

399. Jandric S, Manojlovic S. Quality of Life of Men and Women with Osteoarthritis of the Hip and Arthroplasty Assessment by WOMAC Questionnaire. *American Journal of Physical Medicine & Rehabilitation* 2009;88(4):328-35.

400. 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 and Cartilage* 2012;20 Suppl 1:S99.

401. 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. *Journal of Bone and Joint Surgery-British Volume* 2012;94B(5):648-55.

402. Jolles BM, Michel J, Burnand B, Leyvraz P. Surgical treatment for advanced stage of avascular necrosis of the femoral head in adults. *Cochrane Database of Systematic Reviews* 2006;(3):CD006079.

403. 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(5):538-44.

404. Karas S. Outcomes of birmingham hip resurfacing: a systematic review. *Asian Journal of Sports Medicine* 2012;3(1):1-7.

405. Kenanidis EI, Potoupnis ME, Papavasiliou KA, Sayegh FE, Kapetanos GA. Re: Prospective randomized study of two surgical approaches for total hip arthroplasty. *Journal of Arthroplasty* 2011;26(5):821.

406. Kim S, Losina E, Solomon DH, Wright J, Katz JN. Effectiveness of clinical pathways for total knee and total hip arthroplasty: literature review. *Journal of Arthroplasty* 2003;18(1):69-74.

407. Kim YH, Kim JS, Joo JH, Park JW. Is hydroxyapatite coating necessary to improve survivorship of porous-coated titanium femoral stem? *Journal of Arthroplasty* 2012;27(4):559-63.

408. 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. *Journal of Bone & Joint Surgery - British Volume* 2012;94(6):774-81.

409. 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. *Journal of Bone & Joint Surgery - American Volume* 2002;84-A(8):1372-9.

410. Lane NE. Osteoarthritis of the hip. *New England Journal of Medicine* 2007;357(14):1413-21.

411. 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. *International Orthopaedics* 2007;31(2):199-204.

412. Lavigne M, Masse V, Girard J, Roy AG, Vendittoli PA. [Return to sport after hip resurfacing or total hip arthroplasty: a randomized study]. *Revue de chirurgie orthopédique et réparatrice de l'appareil moteur* 2008;94(4):361-7.

413. 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.

414. 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. *International Orthopaedics* 2011;35(8):1139-43.

415. 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. *Journal of Clinical Nursing* 2008;17(1):40-5.

416. Moskal JT, Capps SG. Acetabular Component Positioning in Total Hip Arthroplasty: An Evidence-Based Analysis. *Journal of Arthroplasty* 2011;26(8):1432-7.

417. 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. *Orthopaedics & traumatology, surgery & research* 2011;97(1):14-21.

418. Naal FD, Impellizzeri FM. How active are patients undergoing total joint arthroplasty? A systematic review. *Clinical Orthopaedics and Related Research* 2010;468(7):1891-904.

419. Nantel J, Termoz N, Vendittoli PA, Lavigne M, Prince F. Gait Patterns After Total Hip Arthroplasty and Surface Replacement Arthroplasty. *Archives of Physical Medicine and Rehabilitation* 2009;90(3):463-9.

420. 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). *Journal of Bone & Joint Surgery - American Volume* 2012;94(10):874-80.

421. 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(2):129-34.

422. 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(3):402-12.

423. 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. *European Cells & Materials* 2004;8:65-73.

424. 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. *Archives of Orthopaedic & Trauma Surgery* 2004;124(7):489-94.

425. Palm L, Jacobsson S-A, Ivarsson I. Hydroxyapatite coating improves 8- to 10-year performance of the Link RS cementless femoral stem. *Journal of Arthroplasty* 2002;17(2):2002.

426. Patel D, Parvizi J, Sharkey PF. Alternative bearing surface options for revision total hip arthroplasty. *Instructional Course Lectures* 2011;60:257-67.

427. Petersen MK, Andersen NT, Soballe K. Self-reported functional outcome after primary total hip replacement treated with two different periopera-tive regimes: a follow-up study involving 61 patients. *Acta Orthop* 2008;79(2):160-7.

428. 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. *Journal of Bone & Joint Surgery - American Volume* 2002;84-A(1):39-48.

429. 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 Review of Medical Devices* 2012;9(3):257-62.

430. Prudhon JL. Dual-mobility cup and cemented femoral component: 6 year follow-up results. *Hip International* 2011;21(6):713-7.

431. 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(1):108-15.

432. 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. *Journal of Arthroplasty* 2004;19(7 Suppl 2):2-9.

433. Ratko TA, Aronson N, Ziegler KM, Bonnell CJ. Metal-on-metal total hip resurfacing. Blue Cross and Clue Cross Shield, Technology Evaluation Center; 2007.

434. 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 Musculoskeletal Disorders* 2008;9:147.

435. 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 & research* 2008;59(6):876-83.

436. 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. *Journal of Arthroplasty* 2004;19(8):962-71.

437. 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. *Canadian Journal of Surgery* 2008;51(6):428-36.

438. 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. *Archives of Orthopaedic & Trauma Surgery* 2006;126(5):324-9.

439. Schmidutz F, Dull T, Voges O, Grupp T, Muller P, Jansson V. Secondary cement injection technique reduces pulmonary embolism in total hip arthroplasty. *International Orthopaedics* 2012;36(8):1575-81.

440. Scott D. Osteoarthritis of the hip. *Clinical Evidence* 2009.

441. 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. *Journal of Bone & Joint Surgery - American Volume* 2006;88 Suppl 3:116-25.

442. Sharma V, Morgan PM, Cheng EY. Factors influencing early rehabilitation after tha: A systematic review. *Clinical Orthopaedics and Related Research* 2009;467(6):1400-11.

443. Shetty V, Shitole B, Shetty G, Thakur H, Bhandari M. Optimal bearing surfaces for total hip replacement in the young patient: a meta-analysis. *International Orthopaedics* 2011;35(9):1281-7.

444. Singh JA, Kundukulam J, Riddle DL, Strand V, Tugwell P. Early postoperative mortality following joint arthroplasty: a systematic review. *Journal of Rheumatology* 2011;38(7):1507-13.

445. Singh JA. Epidemiology of knee and hip arthroplasty: a systematic review. *The open orthopaedics journal* 2011;5:80-5.

446. 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. *Journal of Arthroplasty* 2006;21(3):344-52.

447. Stanat SJC, Capozzi JD. Squeaking in Third- and Fourth-Generation Ceramic-on-Ceramic Total Hip Arthroplasty. Meta-Analysis and Systematic Review. *Journal of Arthroplasty* 2012;27(3):445-53.

448. Stilling M, Rahbek O, Soballe K. Inferior survival of hydroxyapatite versus titanium-coated cups at 15 years. *Clinical Orthopaedics & Related Research* 2009;467(11):2872-9.

449. Suda AJ, Knahr K. Early results with the cementless Variall hip system. *Expert Review of Medical Devices* 2009;6(1):21-5.

450. Tarasevicius S, Robertsson O, Wingstrand H. Posterior soft tissue repair in total hip arthroplasty: A randomized controlled trial. *Orthopedics* 2010;33(12).

451. 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 International* 2012;22(1):41-9.

452. 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. *Journal of Arthroplasty* 2012;27(1):134-42.

453. Thien TM, Thanner J, Karrholm J. Randomized comparison between 3 surface treatments of a single anteverted stem design: 84 hips followed for 5 years. *Journal of Arthroplasty* 2010;25(3):437-44.

454. Thomas W, Tafuro L, Thomas S. Osteoinductive gel in cementless hip joint replacement: a randomized prospective study. *Current Orthopaedic Practice* 2009;20(6):655-9.

455. Timperley AJ, Whitehouse SL, Hourigan PG. The influence of a suction device on fixation of a cemented cup using RSA. *Clinical Orthopaedics & Related Research* 2009;467(3):792-8.

456. 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 International* 2012;22(2):172-8.

457. 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. *Journal of Arthroplasty* 2003;18(7):95-102.

458. 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 International* 2006;16(1):8-17.

459. 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. *Journal of Bone and Joint Surgery-British Volume* 2011;93B(3):298-306.

460. van Gerwen M, Shaerf DA, Veen RM. Hip resurfacing arthroplasty: A systematic review of functional outcome. *Acta Orthop* 2010;81(6):680-3.

461. Veldstra R, van DA, Kraaneveld EC. Comparing alumina-reduced and conventional surface grit-blasted acetabular cups in primary THA: early results from a randomised clinical trial. *Hip International* 2012;22(3):296-301.

462. 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 International* 2007;17(3):137-42.

463. 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. *Physical Therapy* 2011;91(5):615-29.

464. 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 and Cartilage* 2009;17 Suppl 1:S287-S288.

465. 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. *Journal of Arthroplasty* 2012;27(8):1538-43.

466. Zagra L, Bianchi L, Licari V, Champlon C, Giacometti CR. Gait analysis of THA with different head diameters: A prospective randomized study. *Journal of Orthopaedics and Traumatology* 2011;12(1 Suppl):S149-S150.

467. 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 and Cartilage* 2010;18(4):476-99.

468. 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. *Chinese Journal of Clinical Rehabilitation* 2006;10(13):10.

469. Zwartele RE, Witjes S, Doets HC, Stijnen T, Poll RG. Cementless total hip arthroplasty in rheumatoid arthritis: a systematic review of the literature. *Archives of Orthopaedic & Trauma Surgery* 2012;132(4):535-46.

470. Clement RC, Kamath AF, Derman PB, Garino JP, Lee GC. Bipolar sealing in revision total hip arthroplasty for infection: efficacy and cost analysis. *Journal of Arthroplasty* 2012;27(7):1376-81.

471. 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. *Quality of Life Research* 2004;13(1):15-22.

472. Jansson KA, Granath F. Health-related quality of life (EQ-5D) before and after orthopedic surgery. *Acta Orthop* 2011;82(1):82-9.

473. Larsen K, Hansen TB, Thomsen PB, Christiansen T, Soballe K. Cost-Effectiveness of Accelerated Perioperative Care and Rehabilitation After Total Hip and Knee Arthroplasty. *The Journal of Bone & Joint Surgery* 2009;91(4):761-72.

474. Lemon MA, Hamilton PD, Field RE. Comparing total hip and knee replacement costs. *British Journal of Healthcare Management* 2008;14(3):108-12.

475. 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(2):183-8.

476. Rolfson O, Dahlberg LE, Nilsson J, Malchau H, Garellick G. Variables determining outcome in total hip replacement surgery. *Journal of Bone & Joint Surgery, British Volume* 2009;91-B(2):157-61.

477. 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. *Canadian Journal of Surgery* 2012;55(3):181-90.

478. 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. *Journal of Rheumatology* 2007;34(1):165-71.

479 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. *Journal of Bone & Joint Surgery - American Volume* 1993;75(7):961-75

480 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. *Journal of Bone & Joint Surgery - American Volume* 1997;79(1):53-64.

491 Callaghan JJ, Albright JC, Goetz DD, Olejniczak JP, Johnston RC. Charnley total hip arthroplasty with cement. Minimum twenty-five-year follow-up. *Journal of Bone & Joint Surgery - American Volume* 2000;82(4):487-97.