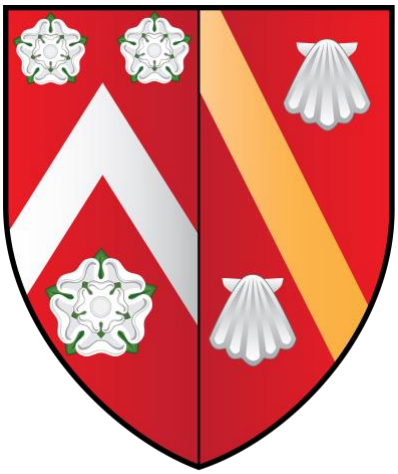
The Investigation of Surrogate Markers for Post-Market Surveillance in Primary Total Hip Arthroplasty

A dissertation submitted to the University of Oxford in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Musculoskeletal Sciences





Tamer T Malak

Wadham College

2021

Abstract

The recent catastrophe of metal-on-metal Total Hip Arthroplasty devices have highlighted the need for early identification of failing devices newly introduced in the market. New regulations have come into force with an emphasis on pre-market clinical evidence generation, transparency, and post-marketing surveillance. However, there are significant limitations with current surveillance using joint registries due to reliance on measuring revision rate. Revision is a rare and typically long-term outcome therefore requiring a significant number of patients undergoing surgery and extended follow-up to detect those devices that are outliers. Therefore, surrogate markers of subsequent revision rate are sought to ensure failing devices are detected early.

The systematic review investigates which available surrogate markers have been validated against the current National Institute for Health and Care Excellence guidelines for an established device. This study found the only validated surrogate marker was implant migration and wear within two-years post-operatively. The only method for measuring wear and migration for both the femoral and acetabular components was Radiostereometric Analysis. However, this has significant limitations in cost, availability, and the failure to universally predict long-term fixation in all devices. Therefore, alternative surrogate markers that are cost-effective are sought.

Chapter 3 investigates analgesic and healthcare utilisation within the first post-operative year as potential surrogate markers of subsequent revision rate in primary

hip arthroplasty surgery. Survival analysis demonstrated increased risk of revision for each surrogate marker. Patients in the top quintile of NSAID use had a sub-hazard ratio (SHR) of 2.20 [95%CI 1.35-3.59; p=0.002], patients using Tramadol SHR 2.42 [95%CI 1.74-3.35; p<0.001], those using atypical analgesics SHR 2.59 [95%CI 1.70-3.95; p<0.001], patients visiting their primary care physician more than four times had a SHR of 2.10 [95%CI 1.29-3.42; p=0.003], patients referred to the Orthopaedic department SHR 1.90 [95%CI 1.16-3.13; p=0.011], and those referred to the Physiotherapy department SHR 2.07 [95%CI 1.29-3.32; p=0.003].

Chapter 4 further investigates whether the novel analgesic and healthcare utilisation surrogate markers can identify differences between device bearings. There was no significant difference in revision rate between metal-on-metal and other bearings in the RACat cohort. However, two surrogate markers were able to identify a difference suggesting that analgesic and healthcare utilisation may potentially be able to identify outliers earlier than revision rate alone.

The findings suggest the possible use of these surrogate markers as part of a surveillance programme to identify potentially failing devices early, based on routinely collected data from registries and, where available, linked electronic medical records. Such surrogate markers can be easily collected at low cost with no impact on patients and physicians.

Funding bodies

Orthopaedic Research UK

Acknowledgements

A special thank you is given to my loving wife Irene for her unconditional support during the thesis and throughout my career. She has been a rock during the difficult times and a beacon during the joyous. I thank both my son, Joshua, and daughter, Olivia, for their never-ending love and giving me joy every day.

A very special thank you to my parents whose support and love throughout my life has made me the man I am today and without whom none of what I achieve would be possible. They have shown me how to live with God as the centre of my life. I am dedicating this thesis to my late father who after a yearlong battle with leukaemia died during my time conducting this research. He has been and continues to be an inspirational role model and has made me the man I am today. Thank you to my mother who has ensured no less support and guidance since.

A sincerest thank you is offered to both of my supervisors, Professor Andrew Judge and Professor Daniel Prieto-Alhambra, for their generous time and insightful support to help guide me during the DPhil. Their mentorship has greatly influenced my professional career and without their guidance I would undoubtedly have not completed this thesis and grown as a researcher.

I would like to give my sincerest thanks to both Professor Afsie Sabokbar and Mr Sam Burnell who have been invaluable at helping me through the rough patches during this process and without whom I would not have achieved this thesis.

I would like to thank Orthopaedic Research UK who have generously supported this DPhil and have been very accommodating through special circumstances.

Abbreviations

AE	Adverse Events
ALVAL	Aseptic Lymphocytic Vasculitis Associated Lesions
AQuAS	Agency for Health Quality and Assessment of Catalonia
ARMD	Adverse Reactions to Metal Debris
ASR	Articular Surface Replacement
ATC	Anatomical Therapeutic Chemical Classification System
AUC	Area Under Curve
BMI	Body Mass Index
BMJ	British Medical Journal
BNF	British National Formulary
BOA	British Orthopaedic Association
CatSalut	Catalan Health Service
CE	Conformité Européenne
CEIC	Comite Etic d'Investigacions Cliniques in Catalan language
CI	Confidence Intervals
CoC	Ceramic-on-Ceramic
CoP	Ceramic-on-Polyethylene
COPD	Chronic Obstructive Pulmonary Disease
CPRD	Clinical Practice Research Datalink
DDD	Daily Defined Doses
eBRA	einzel-Bild-Röntgen-Analyse
EMR	Electronic Medical Records
EPRD	Endo-Prothesen Register Deutschland
EU	European Union

EUDAMED	European Database on Medical Devices
FDA	Food and Drug Agency
F&G	Fine & Gray
GP	General Practitioner
GPRD	General Practice Research Database
GSPR	General Safety and Performance Requirements
HBF	Heterotrophic Bone Formation
HES	Hospital Episode Statistics
HR	Hazard Ratio
IDEAL	Idea, Development, Exploration, Assessment, Long-term Follow-up
IVDR	In-Vitro Diagnostic Device Regulation
KM	Kaplan-Meier
MAR	Missing At Random
MARS	Metal Artefact Reduction Sequence
MCAR	Missing Completely At Random
MCE	Monte Carlo Errors
MDR	Medical Device Regulation
MHRA	Medicines and Healthcare products Regulatory Agency
MI	Multiple Imputation
MNAR	Missing Not At Random
MICE	Multiple Imputation by Chained Equations
MoM	Metal-on-Metal
MoP	Metal-on-Polyethylene
MRI	Magnetic Resonance Imaging
NHS	National Health Service
NICE	National Institute for Health and Care Excellence

NJR	National Joint Registry of England, Wales, Northern Ireland and the Isle of Man
NPV	Negative Predictive Value
NSAID(s)	Non-Steroidal Anti-Inflammatory Drugs
OA	Osteoarthritis
ODEP	Orthopaedic Data Evaluation Panel
OHS	Oxford Hip Score
OVID	Embase database
PIP	Poly Implant Prothèse
PMCF	Post-Market Clinical Follow-up
PMS	Post-Market Surveillance
PPV	Positive Predictive Value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PROM(s)	Patient Reported Outcome Measures
RA	Rheumatoid Arthritis
RACat	Catalan Arthroplasty Registry
RCS	Royal College of Surgeons
RCT	Randomised Controlled Trial
ROC	Receiver Operator Curve
RSA	Radiostereometric Analysis
RWE	Real World Evidence
SCCOT	Society of Orthopaedic Surgery and Traumatology
SHR	Sub-Hazard Ratio
SIDIAP	Sistema Informació Desenvolupament Investigació Atenció Primària
SNRI	Serotonin-Norepinephrine Reuptake Inhibitors
SSCP	European Database on Medical Devices
TCA	Tricyclic Antidepressants

THA	Total Hip Arthroplasty
TIA	Transient Ischaemic Attack
UDI	Unique Device Identification
UK	United Kingdom
WHO	World Health Organisation

Contents

ABSTRACT ..... **II**

FUNDING BODIES ..... **IV**

ACKNOWLEDGEMENTS ..... **V**

ABBREVIATIONS ..... **VI**

CONTENTS ..... **X**

TABLES ..... **XV**

FIGURES ..... **XXI**

1. INTRODUCTION AND LITERATURE REVIEW§ **1**

1.1. SUMMARY 1

1.2. INTRODUCTION 2

1.3. EVOLUTION OF HIP REPLACEMENT DEVICES 4

1.3.1. Cemented devices *4*

1.3.2. Cementless devices *5*

1.3.3. Bearing surfaces *6*

1.4. METAL-ON-METAL BEARING SURFACES 7

1.5. DEVICE REGULATION 13

1.6. EVIDENCE-BASE FOR HIP ARTHROPLASTY DEVICES 23

1.7. CONCLUSION 26

1.8. SCOPE OF THESIS 27

1.9. AIM OF THESIS 27

2. SYSTEMATIC REVIEW OF CURRENT SURROGATE MARKERS OF LONG-TERM OUTCOME IN PRIMARY TOTAL HIP ARTHROPLASTY§ **28**

2.1. SUMMARY 28

2.2. INTRODUCTION 30
2.3. METHODS 32
2.3.1. Objective 32
2.3.2. Eligibility criteria 32
2.3.3. Search strategy 34
2.3.4. Statistical analysis 36
2.4. RESULTS 37
2.4.1. Outcome of search strategy 37
2.4.2. Implant migration and wear 41
2.4.3. Other surrogate outcome markers identified 43
2.4.4. Functional outcome score 43
2.5. DISCUSSION 43
2.5.1. Limitations 49
2.5.2. Clinical relevance 50
2.5.3. Conclusion 52
3. INVESTIGATION OF NOVEL SURROGATE MARKERS OF MID- TO LONG-TERM OUTCOME IN PRIMARY TOTAL HIP ARTHROPLASTY 53
3.1. SUMMARY 53
3.2. INTRODUCTION 53
3.2.1. New device regulations 53
3.2.2. Limitations of registries 60
3.2.3. Pharmaceutical versus medical devices post-market surveillance 64
3.2.4. Potential alternative surrogate markers 67
3.2.5. Analgesic utilisation within the first post-operative year following Total Hip Arthroplasty as a surrogate marker of device survival 70
3.2.6. Healthcare utilisation within the first post-operative year following Total Hip Arthroplasty as a surrogate marker of device survival 80
3.3. METHODS 86

3.3.1. Aim ... *86*

3.3.2. Study design ... *86*

3.3.3. Data source ... *87*

3.3.4. Ethics ... *89*

3.3.5. Study population ... *89*

3.3.6. Outcomes ... *90*

3.3.7. Exposures ... *91*

3.3.8. Confounders ... *104*

3.3.9. Statistical analysis ... *124*

3.4. RESULTS ... 130

3.4.1. Confounders ... *130*

3.4.2. Multiple imputation ... *138*

3.4.3. Study population and baseline characteristics ... *148*

3.4.4. Follow-up ... *150*

3.4.5. Rates of revision ... *151*

3.4.6. Reasons for revision surgery ... *160*

3.4.7. Exposure versus mortality ... *160*

3.4.8. Main outcome ... *162*

3.4.9. Post-estimation analysis ... *165*

3.4.10. Evaluation of multiple imputation ... *180*

3.4.11. Sensitivity analyses ... *181*

3.4.1. Combined analgesic and healthcare utilisation ... *184*

3.5. DISCUSSION ... 186

3.5.1. Limitations ... *197*

3.5.2. Strengths ... *205*

3.6. CONCLUSION ... 206

4. INVESTIGATION OF NOVEL SURROGATE MARKERS OF TOTAL HIP ARTHROPLASTY OUTCOME TO DETECT DIFFERENCES BETWEEN DEVICE BEARINGS ... 207

4.1. SUMMARY 207
4.2. INTRODUCTION 207
4.3. METHODS 208
4.3.1. Aims 208
4.3.2. Study design 208
4.3.3. Outcome 209
4.3.4. Exposures 210
4.3.5. Confounders 210
4.3.6. Statistical Analysis 211
4.4. RESULTS 213
4.4.1. Study population and baseline characteristics 213
4.4.2. Numbers needed to screen 216
4.4.3. Rates of revision 218
4.4.4. Device bearing versus mortality 219
4.4.5. Device bearing versus revision 220
4.4.6. Sample size 220
4.4.7. Device bearing versus novel surrogate markers 221
4.5. DISCUSSION 223
4.5.1. Limitations 227
4.5.2. Strengths 229
4.5.3. Conclusion 229
5. DISCUSSION 230
5.1. SUMMARY OF RESULTS 230
5.2. CLINICAL RELEVANCE 235
5.3. GENERALISABILITY 243
5.3.1. Joint registry 243
5.3.2. Healthcare system 249
5.3.3. Primary care database 251

5.4. LIMITATIONS 251
5.4.1. Limitations of database 251
5.4.2. Limitations of main analysis 253
5.4.3. Limitations of multiple imputation 256
5.5. STRENGTHS 257
5.6. FUTURE WORK 258
5.7. CONCLUSION 262
6. BIBLIOGRAPHY 265
7. APPENDICIES 363
7.1. APPENDIX 1 363
7.2. APPENDIX 2 364
7.3. APPENDIX 3 379
7.4. APPENDIX 4 387
7.4.1. Atypical analgesic multiple imputation results 387
7.5. APPENDIX 5 391
7.5.1. Baseline characteristics for exposures 391
7.6. APPENDIX 6 398

Tables

Table 1-1: The 5-step process for the phased introduction of new surgical devices. 20

Table 2-1: Search strategies employed for Medline, Embase and Cochrane databases. Edited with permission from BJR. 34

Table 2-2: Characteristics of included studies. Edited with permission from BJR. 39

Table 2-3: Validated surrogate markers of long-term revision rate in primary total hip arthroplasty. 41

Table 3-1: Comparison of confounders with outcome variable (revision). 133

Table 3-2: Comparison of adjusted confounders with outcome variable (revision). 136

Table 3-3: Characteristics of device related risk factors. 137

Table 3-4: Proportion of missingness for confounding variables with missing data. ...138

Table 3-5: Baseline characteristics of confounding variables with complete case analysis or all eligible patients in the study. 139

Table 3-6: Auxiliary variables showing statistically significant association with confounding variable containing missing data. 142

Table 3-7: Proportion of patients with a BMI, alcohol, or smoking values and those with missing values for each potential confounder explanatory variable. 143

Table 3-8: Proportion of patients with a BMI, alcohol, or smoking values and those with missing values for each potential exposure explanatory variable. 145

Table 3-9: Proportion of patients with a BMI, alcohol, or smoking values and those with missing values for the potential outcome explanatory variable. 145

Table 3-10: Backward stepwise regression against missingness of BMI, alcohol, and smoking variables. 146

Table 3-11: pre-and post-multiple imputation baseline characteristics for BMI, smoking and alcohol variables. 146

Table 3-12: Baseline characteristics following multiple imputation. 149

Table 3-13: Lifetable of the outcome of interest (revision). 150

Table 3-14: Unadjusted rates of revision for the NSAID exposure. 151

Table 3-15: Unadjusted rates of revision for the Tramadol exposure 153

Table 3-16: Unadjusted rates of revision for the Atypical Analgesic exposure. 154

Table 3-17: Unadjusted rates of revision for the frequency of GP visits exposure. 156

Table 3-18: Unadjusted rates of revision for the referral to the Orthopaedic department exposure. 157

Table 3-19: Unadjusted rates of revision for the referral to the Physiotherapy department exposure. 159

Table 3-20: Reasons for revision surgery. 160

Table 3-21: Univariate and multivariate Cox proportional hazards regression for the analgesic and healthcare utilisation exposures against mortality. 161

Table 3-22: Univariate and multivariate Fine and Gray (F&G) analysis to investigate the association with analgesic and healthcare utilisation within the first post-operative year and subsequent revision rate. 163

Table 3-23: Comparison of Cox proportional hazard regression and Fine and Gray (F&G) analysis of the association between analgesic and healthcare utilisation and subsequent revision rate. 164

Table 3-24: Sensitivity and specificity table with comparisons between the top quintile of NSAID use and other quintiles combined within the first post-operative year. 166

Table 3-25: Sensitivity and specificity table with comparisons between Tramadol users and non-users within the first post-operative year. 169

Table 3-26: Sensitivity and specificity table with comparisons between Atypical Analgesic users and non-users within the first post-operative year. 171

Table 3-27: Sensitivity and specificity table with comparisons between patients with more than four GP visits and those with zero to four within the first post-operative year. 174

Table 3-28: Sensitivity and specificity table with comparisons between patients referred to the Orthopaedic department and those who were not referred within the first post-operative year. 176

Table 3-29: Sensitivity and specificity table with comparisons between patients referred to the Physiotherapy department and those who were not referred within the first post-operative year. 179

Table 3-30: Proportion of patients undergoing revision surgery categorised by sensitivity analysis variable. 181

Table 3-31: Multivariate Fine and Gray (F&G) analysis to investigate the association with analgesic and healthcare utilisation within the first post-operative year and subsequent revision rate comparing the inclusion or exclusion of Charlson score. 183

Table 3-32: Odds ratios following logistic regression for potential surrogate markers combined on a multiple imputation dataset. 184

Table 4-1: Baseline characteristics for THA bearings. 215

Table 4-2: Characteristics for THA bearings against novel surrogate markers 216

Table 4-3: Survivor function for each surrogate marker at the maximum follow up time interval. 217

Table 4-4: Number needed to screen for each surrogate marker 217

Table 4-5: Unadjusted rates of revision for MoM versus other bearing devices. 218

Table 4-6: Univariate and multivariate Cox proportional hazards regression for device bearing against mortality. 219

Table 4-7: Univariate and multivariate Cox proportional regression analysis to investigate the association with device bearing and subsequent revision rate. 220

Table 4-8: Estimated sample size for each analgesic and healthcare utilisation surrogate marker. 221

Table 4-9: Univariate and multivariate logistic regression analysis to investigate rates of novel surrogate maker between metal-on-metal bearings versus other bearings. 222

Table 7-1: Rubin types of missing data 366

Table 7-2: Number of imputations required to reduce loss of power according to proportion of missingness. Adaptation from Graham et al. 376

Table 7-3: Proportion of missingness for confounding variables with missing data for the Atypical Analgesic cohort. 387

Table 7-4: Baseline characteristics of confounding variables with complete case analysis or all eligible patients in the study for the Atypical Analgesic cohort. 387

Table 7-5: pre-and post-multiple imputation baseline characteristics for BMI, smoking and alcohol variables for the Atypical Analgesic cohort. 389

Table 7-6: Baseline characteristics for the NSAID exposure. 391

Table 7-7: Baseline characteristics for the Tramadol exposure. 393

Table 7-8: Baseline characteristics for the Atypical Analgesic exposure. 394

Table 7-9: Baseline characteristics for the frequency of GP visits exposure. 395

Table 7-10: Baseline characteristics for the referral to the Orthopaedic department exposure. 396

Table 7-11: Baseline characteristics for the referral to the Physiotherapy department exposure. 397

Table 7-12: Comparison of pre- and post-multiple imputation univariate regression for the NSAID exposure in both Cox proportional hazard and Fine & Gray models.398

Table 7-13: Multiple imputation Monte Carlo errors for the NSAID exposure. 399

Table 7-14: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the NSAID exposure.....400

Table 7-15: Comparison of pre- and post-multiple imputation univariate regression for the Tramadol exposure in both Cox proportional hazard and Fine & Gray models. 400

Table 7-16: Multiple imputation Monte Carlo errors for the Tramadol exposure. 401

Table 7-17: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for Tramadol exposure.....401

Table 7-18: Comparison of pre- and post-multiple imputation univariate regression for the Atypical Analgesic exposure in both Cox proportional hazard and Fine & Gray models.....402

Table 7-19: Multiple imputation Monte Carlo errors for the Atypical Analgesic exposure. 402

Table 7-20: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the Atypical Analgesic exposure. 402

Table 7-21: Comparison of pre- and post-multiple imputation univariate regression for the frequency of GP visits exposure in both Cox proportional hazard and Fine & Gray models.....403

Table 7-22: Multiple imputation Monte Carlo errors for the frequency of GP visits exposure. 404

Table 7-23: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the frequency of GP visits exposure.....405

Table 7-24: Comparison of pre- and post-multiple imputation univariate regression for the referral to the Orthopaedic department exposure in both Cox proportional hazard and Fine & Gray models. 405

Table 7-25: Multiple imputation Monte Carlo errors for the referral to the Orthopaedic department exposure. 406

Table 7-26: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the referral to the Orthopaedic department exposure. 406

Table 7-27: Comparison of pre- and post-multiple imputation univariate regression for the referral to the Physiotherapy department exposure in both Cox proportional hazard and Fine & Gray models. 407

Table 7-28: Multiple imputation Monte Carlo errors for the referral to the Physiotherapy department exposure. 407

Table 7-29: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the referral to the Physiotherapy department exposure. 408

Figures

Figure 1-1: Illustrations of common hip devices (i) Charnley stem (cementless). (ii) Exeter stem (cemented). (iii) Birmingham Hip Resurfacing System. Edited with permission from ARUK.....3

Figure 1-2: (i) MRI depicting a pseudotumour with local soft tissue and bone destruction. (ii) Intra-operative view of revision surgery for a metal-on-metal hip replacement showing fluid collection (brown). Edited with permission from ARUK.....9

Figure 1-3: ODEP ratings at the time of MoM recalls. (i) For devices meeting the NICE benchmark of 10 years follow-up. (ii) For devices with evidence less than the NICE benchmark of 10 years but with evidence at 3, 5 or 7 years. Edited with permission from ARUK. 19

Figure 2-1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow-chart. Edited with permission from BJR.....38

Figure 3-1: Sources of evidence for clinical evaluation of medical devices.....56

Figure 3-2: Histogram of NSAID exposure measured as Daily Defined Dose (DDD). ..92

Figure 3-3: Transformations of NSAID exposure measured as Daily Defined Dose (DDD).....93

Figure 3-4: Histogram of Tramadol exposure measured as Daily Defined Dose (DDD). 94

Figure 3-5: Transformations of Tramadol exposure measured as Daily Defined Dose (DDD).....95

Figure 3-6: Histogram of Atypical Analgesic exposure measured as Daily Defined Dose (DDD).....97

Figure 3-7: Transformations of Atypical Analgesic exposure measured as Daily Defined Dose (DDD). 98

Figure 3-8: Histogram of frequency of GP visits exposure measured as a discrete variable. 99

Figure 3-9: Transformations of frequency of GP visits exposure measured as a discrete variable. 100

Figure 3-10: Fractional polynomial graph for frequency of GP visits against the outcome of interest (revision) 101

Figure 3-11: Histogram of age variable. 130

Figure 3-12: Transformations of age variable. 131

Figure 3-13: Histogram of body mass index variable. 132

Figure 3-14: Distribution of missing variables following multiple imputation. i) BMI histogram within combined datasets. ii) BMI histogram within each dataset. iii) Smoking distribution within each dataset. iv) Alcohol distribution within each dataset. 147

Figure 3-15: Patient selection flow-chart 148

Figure 3-16: Kaplan-Meier survival estimates for the NSAID exposure. 152

Figure 3-17: Kaplan-Meier survival estimates for the Tramadol exposure. 153

Figure 3-18: Kaplan-Meier survival estimates for the Atypical Analgesic exposure. 155

Figure 3-19: Kaplan-Meier survival estimates for the frequency of GP visits exposure. 156

Figure 3-20: Kaplan-Meier survival estimates for the referral to the Orthopaedic department exposure. 158

Figure 3-21: Kaplan-Meier survival estimates for the referral to the Physiotherapy department exposure. 159

Figure 3-22: Receiver operator curve (ROC) and area under the ROC curve for the NSAID exposure. 165

Figure 3-23: Cumulative incidence function for the NSAID exposure comparing the top quintile and other quintiles combined. 167

Figure 3-24: Receiver operator curve (ROC) and area under the ROC curve for the Tramadol exposure. 168

Figure 3-25: Cumulative incidence function for the Tramadol exposure. 170

Figure 3-26: Receiver operator curve (ROC) and area under the ROC curve for the Atypical Analgesic exposure. 171

Figure 3-27: Cumulative incidence function for the Atypical Analgesic exposure. 172

Figure 3-28: Receiver operator curve (ROC) and area under the ROC curve for the frequency of GP visits exposure. 173

Figure 3-29: Cumulative incidence function for the frequency of GP visits exposure. 175

Figure 3-30: Receiver operator curve (ROC) and area under the ROC curve for the referral to the Orthopaedic department exposure. 176

Figure 3-31: Cumulative incidence function for the referral to the Orthopaedic department exposure. 177

Figure 3-32: Receiver operator curve (ROC) and area under the ROC curve for the referral to the Physiotherapy department exposure. 178

Figure 3-33: Cumulative incidence function for the referral to the Physiotherapy department exposure. 180

Figure 3-34: Receiver operator curve (ROC) and area under the ROC curve for all exposures combined. 185

Figure 4-1: Metal-on-Metal bearing analysis patient selection flow-chart. 214

Figure 4-2: Kaplan-Meier survival estimates for device bearing. 219

Figure 7-1: Distribution of missing variables following multiple imputation for the Atypical Analgesic cohort. i) BMI histogram within combined datasets. ii) BMI histogram within each dataset. iii) Smoking distribution within each dataset. iv) Alcohol distribution within each dataset. 390

1. Introduction and Literature Review[§]

1.1. Summary

Total Hip Arthroplasty is hailed as one of the best innovations of the 20th Century. However, there have been several devices with catastrophic failings, most recently metal-on-metal devices. There have been calls to urgently revise current regulations surrounding medical devices to avoid such failings in the future. The need for a phased introduction of new devices with robust post-market surveillance has been highlighted. Currently, there are excellent resources that could be used to potentially identify failing implants early, but have several limitations. The number of available implants has increased, but with deficiencies in the evidence-base demonstrating clinical effectiveness. However, current resources are not adequate at identifying failing devices newly introduced into the market. Therefore, alternative surrogate markers are sought.

[§]Work described in this chapter has been published as[1]:

Malak TT, Beard D, Glyn-Jones S. Total Hip Arthroplasty: Recent Advances and Controversies. Arthritis Research UK Topical Reviews 2014; Spring 2014 (Issue 4 Series 7). url: http://www.arthritisresearchuk.org/health-professionals-and-students/reports/topical-reviews/topical-reviews-spring-2014.aspx. Re-use permitted by ARUK. The text includes minor and major additions, and formatting changes to the original text. Figures or Tables may have been reproduced with edits.

Kynaston-Pearson F, Ashmore AM, Malak TT, et al. Primary hip replacement prostheses and their evidence base: systematic review of literature. BMJ 2013; 347: f6956. Re-use permitted by BMJ. The text includes minor and major additions, and formatting changes to the original text. Figures or Tables may have been reproduced with edits.

1.2. Introduction

Total Hip Arthroplasty (THA) is considered one of the most successful and cost-effective procedures in modern medicine[2]. THA was introduced and developed in the 1960s by the British Orthopaedic surgeon Sir John Charnley[3]. His design of prosthesis and fundamental principles of hip replacement surgery are still used today. For patients suffering with hip pain and functional deficit caused by a variety of diseases, THA is effective at relieving pain, restoring function and improving a patient's overall quality of life[4–9]. The World Health Organization (WHO) considers hip arthroplasty to be one of the most cost-effective and successful interventions in modern medicine to tackle the burden of musculoskeletal disease[10]. The latest report from the National Joint Registry (NJR) of England and Wales reports an average of over 90,000 THAs are performed per year[11,12]. Latest figures from the NJR reports have shown a nearly 20% rise between 2007 and 2017[13]. Several other studies have predicted significant rises to occur with estimates of 70%[14], 134%[15], 174%[16], and 276%[17] by 2030 across the world. The increasing number of obese individuals and an ageing population determined to remain active are likely to drive this trend further[12,18].

THA surgery aims to alleviate pain and improve function by replacing diseased hip articular surfaces with inert and durable synthetic material. Figure 1-1 shows illustrations of common types of hip implants used in THA.

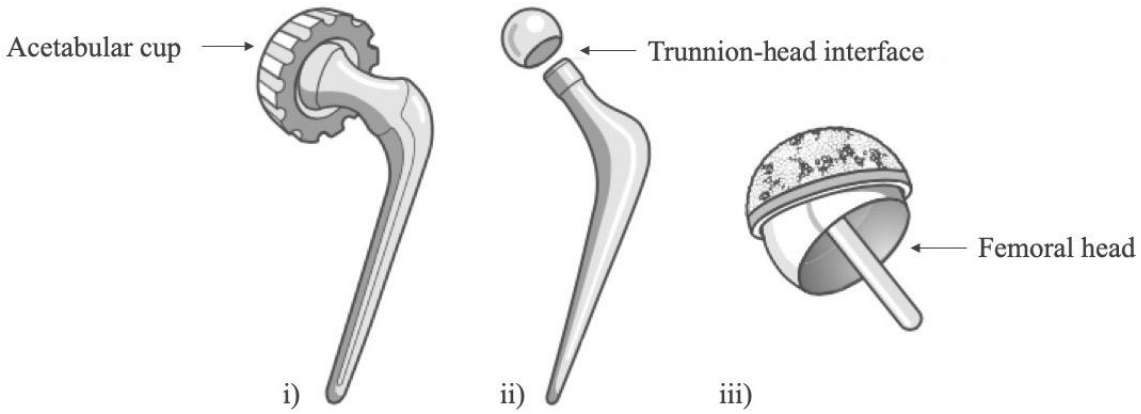


Figure 1-1: Illustrations of common hip devices (i) Charnley stem (cementless). (ii) Exeter stem (cemented). (iii) Birmingham Hip Resurfacing System. Edited with permission from ARUK.

Despite being considered a cost-effective treatment, THA is primarily reserved as a last resort option once all non-operative approaches, such as lifestyle modification and pharmacological treatment for pain control, have been exhausted. By significantly improving pain and function, people with severe hip disease, can see a dramatic change in their quality of life, making THA a potentially life-changing procedure[4–9]. There is extensive clinical evidence showing that THA has excellent long-term survivorship in both the younger and older patient groups[19–21]. In the case of the conventional prosthesis, such as the cemented Charnley implant, survival rate is typically greater than 90% at ten years[22]. Furthermore, expected THA survival can be achieved at rates of 85% and 78% at 15 and 20 years respectively[23]. There are studies investigating 25-year survivorship reporting rates over 70%[22,23].

THA surgery has substantial evidence to support its claim as one of the most important advancements in modern medicine, particularly considering excellent outcomes. However, there have been areas of concern. Newer designs of THA have been extensively reported in the media following recent studies and the NJR showing higher

revision rates. This has ignited the debate of medical device regulation in the UK, Europe, and North America. This chapter describes these issues in relation to the current evidence base.

1.3. Evolution of hip replacement devices

The first prosthetic total hip replacement was developed by Wiles in the 1930s and is held as the beginning of the modern age for THA[24]. It was Charnley who revolutionised the procedure introducing the key principles of low friction, cement use, and high-density polyethylene for the bearing[25]. At this time hip replacement was reserved for the elderly and infirm patients. However, indications for surgery evolved to include compromises in quality of life, with patients wishing to stay active. Innovations occurred to develop high-performance devices to meet patient expectations and ensure decreased failure rates. Thus, new materials and bearing surfaces were developed.

1.3.1. Cemented devices

Sir John Charnley introduced the concept of cement in the 1950s to be used for the fixation of hip devices[26]. From this technique, longevity of hip implants was possible due to the durability of fixation at the bone-cement interface. One of the important determining factors of this technique is the cement as it is manufactured at the time of surgery. Significant development has been made in how this is performed but the composition of the cement has largely remained the same. Initial techniques required no bone preparation with no pressurisation resulting in poor penetration of the cement

into the bone and thus poor fixation. Modern hip replacement acknowledges that cement acts as a grout rather than a glue. Therefore, increased pressurisation of cement was shown to increase tensile and shear strengths at the bone-cement interface[27]. Today, several techniques are employed to prepare bone for cement including pulsed lavage, retrograde insertion, and pressurisation. These have been shown to increase longevity[28]. There are two design concepts for cemented implants: taper-slip and composite beam, both of which have reported excellent survival rates[29–31]. However, utilising the ideas of one design to another did not always produce satisfactory results. To reduce manufacturing costs the Exeter stem was produced with a matt surface. Unfortunately, this resulted in significantly increased failure rates[32]. Despite being based on the Charnley system, the 3M Capital Cemented Hip System (3M Health Care Limited) had decreased survival due to a thin cement mantle[33]. These examples highlight small changes in design can lead to potential catastrophic problems. Despite the success of cemented implants, issues with early failure were noted due to local inflammatory response by cement particles[34]. Thus, techniques for hip replacement without cement were explored.

1.3.2. Cementless devices

Cementless femoral and acetabular components encourage osseointegration of bone within the implant to provide stability and are largely made with a porous coating. Once stabilisation is achieved, normal transmission of biomechanical forces could be possible. Modern stems have proximal circumferential porous coating to ensure proximal loading and reduce the increased failure and fracture rates of early cementless stems[35]. The shape of cementless stems has been found to be significant.

Anatomic type stems were designed with a curve to mimic the natural bow of femoral bone in the hope of increasing initial stability and subsequent osseointegration. However, despite the theoretical potential of enhanced physiological loading at the femur, higher incidence of thigh pain was found[36]. Cylindrical type stems required larger diameters to achieve canal-filling and subsequent initial stability. Unfortunately, increasing the diameter results in increased incidence of thigh pain[37]. Modern cementless implants are tapered designs using proximal osseointegration and "three-point stem fixation" technique to ensure good initial stability[2]. Excellent outcomes have been reported with thigh pain virtually eliminated[38].

No difference in survivorship has been shown between cobalt chrome or titanium, however, titanium is "notch-sensitive" and therefore more predisposed to cracks[2].

To counteract the difficulties of using cemented acetabular cups and potentially high rates of loosening[39], cementless cups were developed. Initial stability for cementless cups is achieved by press-fitting a slightly larger device into a smaller acetabulum. Subsequent osteointegration occurs into an entirely porous cup. Additional stability is sometimes achieved by using a screw or spikes if the press-fit is not entirely adequate. Good survival rates have been shown for both press-fit and additional screw-fit acetabular devices[40]. However, survival for cementless cups are poorer in comparison to cemented devices, yet there is widespread use of cementless cups[41].

1.3.3. Bearing surfaces

One of the major limitations of conventional total hip arthroplasty remains osteolysis caused by wear particles from polyethylene[42,43]. Wear caused from the harder femoral

head against the softer polyethylene liner causes synovitis, osteolysis, joint instability and ultimately loosening of the implant. Therefore, alternative bearing surfaces are sought with concepts such as hard-on-hard bearings and harder wearing plastic. Examples of hard-on-hard bearings are ceramic-on-ceramic and metal-on-metal. Advances in manufacturing techniques such as ultra-high-molecular-weight polyethylene and gamma irradiation have produced a cross-linked polyethylene. These new bearings have the potential to decrease wear and subsequent osteolysis to increase survivorship. Survival of ultra-cross linked polyethylene is promising when compared to conventional polyethylene[44]. Once again, theoretical advantages and in-vitro studies have not translated into patients, with metal-on-metal devices causing considerable problems.

1.4. Metal-on-metal bearing surfaces

Advancements in THA surgery to reduce complications and improve survivorship includes the development of different bearing types[45]. For younger patients and older patients who are active, conventional bearings can suffer with increased wear and subsequently earlier failure[46,47]. In addition, conventional THAs with small bearing surfaces (22-28mm) can lead to increased dislocation rates in these age groups[48]. An increasing trend to perform replacement hip surgery in the more active or younger patient cohort gave rise to the Metal-on-Metal (MoM) bearing design which can address the complications of increased wear and dislocation risk[18]. Conventional hip replacements use a polyethylene-lined acetabular cup articulating with a metal head[46,49]. As the joint is used, the polyethylene liner wears against the harder femoral head[50]. This gives rise to polyethylene-wear debris which can lead to an inflammatory

reaction and subsequent osteolysis around the joint replacement devices[42,43,51]. In contrast, both the acetabular liner and the femoral head in MoM devices are constructed from cobalt chrome metal, a hard material[46]. The use of this harder material was coupled with manufacturing advances ensuring smooth bearing surfaces that trap a layer of fluid reducing friction[46]. Theoretically, the bearing surfaces do not make contact, and the faster a patient moves the thicker this layer fluid becomes. Combined, these factors produced significantly less wear compared to conventional bearings in *in-vitro* studies[52–54]. A further advantage is cobalt chrome can be manufactured to form very thin components which allows for larger femoral head sizes. The use of large-diameter heads can reduce the risk of dislocation. In addition, thinner components result in more bone conservation leading to greater physiological femoral loading through the hip joint preserving the quality of bone[55], making future revision surgery less complex[56]. This concept of reversibility is one of the main selling points for MoM implants, especially for younger patients, despite little clinical evidence[48]. Therefore, the several theoretical benefits to using MoM technology facilitated the popular and rapid uptake of these devices, particularly in a younger cohort.

MoM hip devices had good initial results with 5-year and long-term follow-up data, and were subsequently widely used following their publication[57,58]. These studies were so called "designer series", using data from patients whose surgery was performed by surgeons involved in the design process of the prosthesis. Prior to concerns being noted over 1,000,000 MoM devices, both resurfacing and stemmed, had been implanted worldwide[59]. However, despite the theoretical advantages and initial good results,

high failure rates with the use of MoM devices were found through independent studies.

The first major report concerning an MoM related issue in a resurfaced hip was made by the Medicines and Healthcare products Regulatory Agency (MHRA) in 2006[60]. Further case-series studies demonstrated failure of MoM devices was often associated with large masses and/or cysts visible on magnetic resonance imaging (MRI) or ultrasound[61]. When revision surgery was performed it was found that these masses were associated with local soft tissue and bone destruction (Figure 1-2).

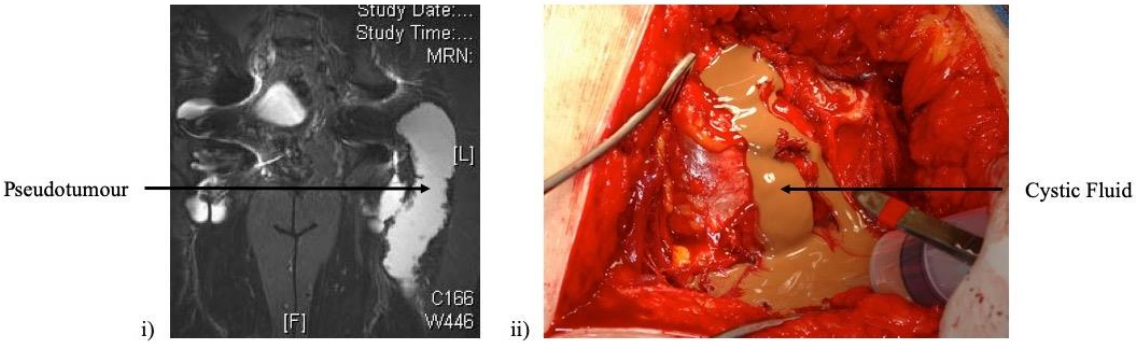


Figure 1-2: (i) MRI depicting a pseudotumour with local soft tissue and bone destruction. (ii) Intra-operative view of revision surgery for a metal-on-metal hip replacement showing fluid collection (brown). Edited with permission from ARUK.

Initially described as 'inflammatory psudotumours'[62] due to their locally destructive nature, it was used as a clinical term to describe a peri-prosthetic mass with evidence of inflammation either clinically, radiologically or histopathological examination[45]. These masses have subsequently had various descriptions including metallosis[63], aseptic lymphocytic vasculitis associated lesions (ALVAL)[64] and adverse reactions to metal

debris (ARMD)[65]. A macrophage dominated inflammatory response associated with large areas of necrosis was found with histological examination[63,66]. In-vitro studies subsequently demonstrated the inflammatory response was initiated by wear debris and corrosion of the cobalt metal which is highly toxic[67,68].

The pathogenesis of failure and associated pseudotumours in MoM devices has been extensively studied. Positioning of the acetabular component is related to the subsequent wear rate and metal ion release[69]. Studies investigating retrieved MoM implants have demonstrated 80% of the pseudotumours associated with increased wear are as a result of edge-loading of the socket, i.e. where there is an unequal load across the acetabular cup leading to load on the rim of the cup[62,65]. There are several factors thought to contribute including implant design, implant positioning and patient factors. It is worth noting that in contrast, 20% of patients who develop pseudotumours have no evidence of increased wear rates. Therefore, one theory is that individuals may be more sensitive to metal wear debris compared to the general population. Currently, these individuals are not easily identifiable, as there is a poor correlation between deep hypersensitivity and cutaneous sensitivity[68].

There are several factors influencing the incidence of MoM induced pseudotumours which varies according to implant, geographical location and series published. One of the most widely publicised MoM implant failings was the resurfacing device, DePuy Articular Surface Replacement® (ASR)[46]. The ASR had a significantly higher revision rate when compared to other designs[70]. It is thought this is due to several design flaws, for example a shallower acetabular cup resulting in greater edge loading and subsequently increased rates of wear compared to other designs. These factors were

highlighted by one of the Birmingham Hip Resurfacing designers, Mr McMinn, citing potential issues as shallowness of the rim, and the manufacturing process used can both lead to increased wear[46]. Furthermore, the design left the implant less forgiving to surgical technique, which should be considered when designing a successful device[46].

Despite initial excellent survival data from designer studies, reporting a cumulative revision rate of less than 0.5%[57], subsequent independent studies discovered higher rates of revision, ranging from 1% to 9.8%[65]. Two risk factors were identified for higher rates of revision secondary to adverse reaction from metal debris: an eightfold increase was seen in patients of female sex, and a threefold increase seen in patients under the age of 40 years[62]. Following independent studies, the NJR confirmed higher revision rates. In the 10th Annual report published in 2013, the NJR reported the nine-year cumulative percentage probability of first revision for any fixation of primary THA, regardless of the bearing type, as 5.06%[71]. MoM devices, irrespective of fixation type, faired significantly worse with cemented, cementless and hybrid implants having a risk of revision of 21.43%, 17.66% and 14.86% respectively[71]. The same report found that not all MoM implants performed equally with variation in performance between differing resurfacing devices. The probability of first revision at eight-years for the Birmingham Hip Resurfacing System was 6.61% compared to the ASR with 29.69%[71]. In addition, it was noted that there were particular issues with larger head stemmed MoM devices which was confirmed in other independent studies[72]. High wear at the trunnion-head interface was also found in these devices[73]. These findings are not isolated to the NJR and have been mirrored in the Australian Joint Registry showing increased revision rates in MoM devices[74].

The long-term effect of these implants is yet to be established with concerns of potential complications for systemic exposure to metal ions[75]. Theoretically, a high metal ion concentration can increase the risk of certain cancers and there have been reports it is associated with cardiac failure[75]. However, there has been no association between MoM devices and subsequent mortality, cancer or heart failure in large cohort studies[76,77]. Early outcomes following revision for MoM devices are poor due to extensive bone destruction and muscle damage[61,78]. MHRA recommended screening to identify patients early in the hope of improving outcomes[79]. Initially, the incidence of pseudotumours secondary to MoM hip replacements being asymptomatic was thought to be low (4%)[68]. However, studies investigating pseudotumour incidence using MRI have described 'silent' pathology in 25% of patients with the best possible Oxford Hip Score[80]. Other studies have shown asymptomatic pseudotumours occurring in 61% of patients[81,82], thus confirming the need for screening. Further to the MHRA recommendations, the British Orthopaedic Association (BOA) have issued guidance for MoM hip replacement follow-up[83]. These guidelines call for all patients to be followed up annually for the life of the implant. Patients with a stemmed component with femoral head diameter >36mm, or any DePuy ASR hip replacement require imaging, either with ultrasound or with MRI using Metal Artifact Reduction Sequence (MARS), and metal ion blood tests. Those patients with hip resurfacing devices or with a stemmed component with femoral head diameter <36mm require imaging only if symptomatic. If blood tests are positive (>7ppb) then a second blood test at 3 months is required. Revision surgery is considered if imaging is abnormal and/or blood metal ion levels are rising. However, follow-up has had significant healthcare cost implications across the world. The annual cost for MoM device follow-up in the UK, as

per the MHRA guidelines, has been estimated at over £8,000,000[84], and almost three-fold for the Food and Drug Administration (FDA) protocol.

The future of MoM devices is not set in stone despite their use being on a steep decline reducing from 15% at the peak of use to less than 1%[83]. Certain devices, such as the ASR, have been withdrawn from the market[46]. The MoM fiasco has been described as one of the "biggest disasters in orthopaedic history"[46], calling into question how these devices were allowed to come to market.

1.5. Device Regulation

Following the public failings of several implants, there has been significant and wide debate concerning current medical device regulation[46,85,86]. Public and professional confidence in the EU systems is at a low point. As discussed, one of the most recent disasters has emerged from the Orthopaedic world with the premature failure of some MoM hip implants[46]. Both the British Medical Journal (BMJ) and the BOA have called for significant change in the regulation of medical devices in Europe[87,88]. Several proposals have been put forward by the European Commission which are currently under negotiations with governments in the EU member states. The BOA has been a drive for change, demanding a good evidence base for all Orthopaedic implants and particularly those newly introduced to the market. More importantly, the BOA have called for radical changes to current medical device regulation and have expressed this to the European Union (EU) in their Practice Strategy document "Restoring your Mobility"[87,88].

Regulation in Europe for medical devices has changed little since its introduction in the 1990s to standardise a set of technical legislation of 'essential requirements' to ease trade between EU member states[89]. However, these requirements do not relate to standards in clinical study design, sample size, selection of endpoints, but focus instead on non-clinical testing[90]. In the USA, medical devices are regulated by a single body, the FDA, by a similar process to pharmacological regulation in Europe through the European Medicines Agency. The American system has come under scrutiny due to inadequate post-approval monitoring of devices. However, medical device regulation in Europe differs substantially to a single agency approach. Under the current European requirements, medical devices only need to successfully apply for a Conformité Européenne (CE) mark to be readily available to surgeons for use in patients[90]. Therefore, multiple agencies are used to evaluate the safety and reliability of medical devices in Europe.

There are approximately 76 separate notified bodies available for manufacturers to achieve a CE mark[90–92], which are appointed by Competent Authorities in each of the EU member states. An example of a Competent Authority in the UK is the MHRA. However, several concerns with the current system have been raised. Firstly, there is potential for variation across Europe in standards and subsequent award of a CE mark, as guidelines ask for non-clinical 'essential requirements' only. Secondly, it is possible for manufacturers to then explore different notified bodies in order to achieve a CE mark for their device, creating a scenario whereby a device is rejected by 75 of the 76 notified bodies but subsequently accepted by the last. Thirdly, evidence submitted to an approving body is not made public and therefore not open to scrutiny[85]. It is still

unclear what the scientific rationale was for the ASR implant for it to be granted approval[46]. Fourthly and perhaps more concerningly, a CE mark may be awarded on the basis of similarity to an existing proven product already approved with a CE mark, without any evidence[92]. The claim of 'equivalency' has been heavily critised[93], and has been highlighted by the recent failings of MoM devices with the ASR devices being introduced into the market having not been tested in humans[46,92]. The notable absence of clinical studies of devices implantable in patients prior to approval has caused concern and has been previously raised over a decade ago following the failure of 3M Capital System[33]. Once a CE mark is achieved, even without clinical evidence, the device can be marketed and used throughout the EU. Joint replacement implants require a higher level of scrutiny since being reclassified in 2007 from a Class II to Class III devices. All class III devices require the submission of human clinical investigations evidence in the application for a CE mark, however, this may not necessarily be new research specific to the device in question. If a manufacturer claims similarity of its device to an existing product then a CE mark can be achieved without new studies[92]. However, even small changes in design can have deleterious effects on implant survival and patient safety as demonstrated in previous devices[94,95]. Despite this, there is limited pre-clinical evidence and post-market surveillance of their performance.

Proposals have been put forward by the European Commission in 2012 and are due to come into effect in 2020 (possible delays due to Coronavirus)[90]. These will address the concerns in current regulation with stronger supervision of Notified Bodies by the Competent Authorities, greater power for Notified Bodies to ensure thorough testing and regular checks, an extended database on medical devices, improved traceability to

assist in effective recalls of failing devices, reinforced rules on the required pre-market clinical data and post-market assessment of devices, and the establishment of a Medical Device Coordination Group comprising representatives of the national Competent Authorities to improve coordination between member states. There are barriers to significant change with some authors arguing that excessive regulation may stifle innovation[96,97]. Despite this a House of Commons committee has investigated this issue in 2012[98], subsequently the Government issued a report outlining its negotiating position[99].

Independently of the Government response, the MHRA initiated a consultation on 'Revision of European Legislation on Medical Devices' to address the issue of regulation. The BOA and Versus Arthritis issued a joint submission to the MHRA consultation, expressing their support for the principles of improving both pre and post-market clinical evaluation[100]. However, they urged caution stating that 'regulatory requirements for clinical investigations do not become overly complex and so a deterrent to those seeking to conduct them'. The BOA subsequently initiated a project with the MHRA named 'Beyond Compliance' with the aim of actively improving the regulation of new devices in the UK to improve safety and quality[101]. Beyond Compliance seeks to complement the existing system by improving "the rigour of processes around CE", "providing guidance and support for the safe and agreed introduction of innovations", and "provide high quality surveillance…to identify failures at the earliest point". It is worth noting, Beyond Compliance is currently a voluntary scheme and does not replace regulatory bodies as EU law prohibits trade barriers between member states.

Fortunately, in the UK, there are existing agencies that ensure quality control, safety monitoring and allow surgeons to compare performance of devices. The first of these is the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man (NJR) collecting data on hip, knee, ankle, elbow and shoulder joint replacements, to monitor their performance[11]. The NJR was founded in 2002 following the failings of the 3M Capital Hip System to aid in the detection of poorly performing new implants, and help identify those patients with the implant for removal[33,102]. Sir John Charnley had recommended a central registry decades before[102]. This is an excellent source of information, with almost universal coverage of all THA procedures performed in the country, and therefore highly representative of all types of patients and health professionals. The aim is to monitor the performance of these prostheses and allow for the identification of patients should a device be recalled. Since April 2011 reporting has become mandatory for hospitals in the National Health Service[103,104]. As a result, the vast majority of operations performed are captured in the NJR database with compliance varying per hospital between 93% and 100% in the NJR annual report of 2012[103].

However, in its current format the NJR does not recognise problems with implants until they are revised, therefore its designed to assess long term outcome, and may fail to detect early failures in new devices[105]. Multiple factors contribute to the potential lack for early detection. One important issue is new implants are typically introduced in small numbers initially, making outliers difficult or impossible to detect[105]. The failures of MoM implants exemplify this issue with example of the ASR device. Before the ASR was noted to have increased revision rates, nearly 100,000 devices worldwide and 10,000 in the UK had been implanted[106]. Moreover, it was single centre cohort studies that identified problems with these devices, several years before the NJR

identified them as outliers[62,65]. Perhaps if compliance was better pre-2011 with more data made available then the failings of MoM devices may have been recognised earlier.

The Orthopaedic Data Evaluation Panel (ODEP), a branch of the NHS supply chain, was formed in 2002 alongside the NJR to set a benchmark for which device performance can be compared[107]. Manufacturers are requested to submit clinical data of their product in the form of cohort studies monitoring failure rates. However, registering with ODEP and submitting data is not compulsory. Nor is it compulsory for a device to receive an ODEP rating for it to be used in the NHS[108]. Devices are rated according to their clinical performance as measured by revision rate. Classification has subsequently changed since the failings of MoM devices; however, it is important to note how implants are rated during this period. The first part of classification is the number of years evidence is available post-implantation. The second part is the level of evidence determined by failure rates. The highest level is Level A for which a device has shown a failure rate of less than 5%. Levels B and C are weaker in turn. Performance is measured at three, five, and ten years. Therefore, the highest possible ODEP rating for a device is 10A indicating low revision rates at ten years and meeting with National Institute of Health and Care Excellence (NICE) guidelines[109] (Figure 1-3).

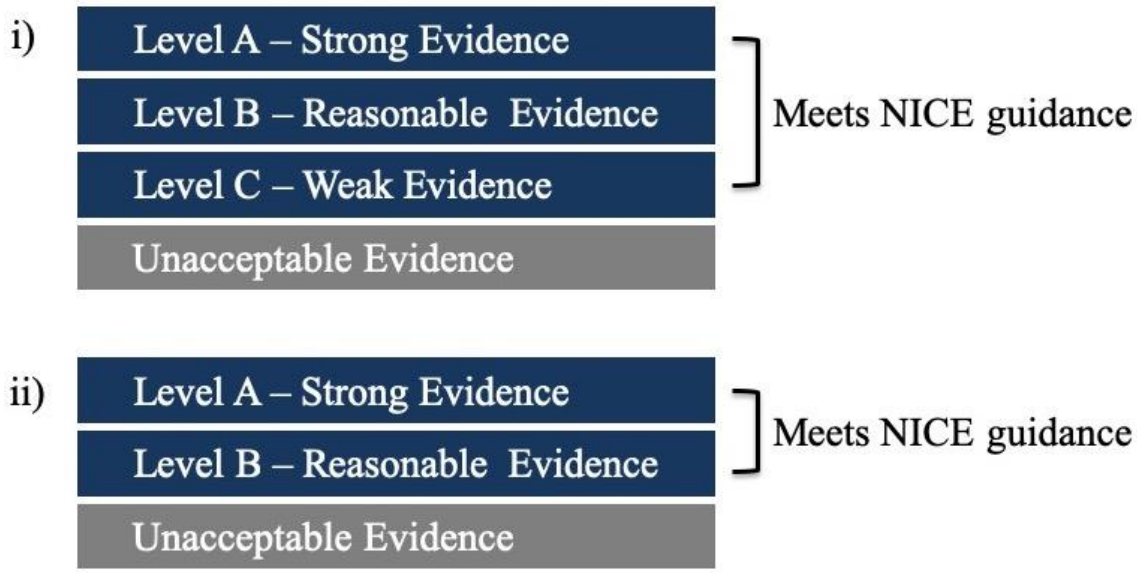


Figure 1-3: ODEP ratings at the time of MoM recalls. (i) For devices meeting the NICE benchmark of 10 years follow-up. (ii) For devices with evidence less than the NICE benchmark of 10 years but with evidence at 3, 5 or 7 years. Edited with permission from ARUK.

NICE have set a benchmark of a ten-year ODEP rating with a revision rate of less than 5% for a proven device[109]. Any implant falling short of this, including devices newly introduced on the market, should only be used as part of ongoing clinical trials. A device lacking evidence that meets the three-year mark may be classified as 'pre-entry' on the condition that manufacturers update ODEP with post-market surveillance data. It is important to note that inclusion as 'pre-entry' does not require manufacturers to submit evidence of use from peer-reviewed publications, whether from pre-clinical or clinical stages of development. Devices that meet none of the criteria thus far and have no evidence submitted by manufactures are termed 'unclassified'. ODEP is an excellent system to determine evidence base of devices and be a marker of implant quality. Despite the potential benefit to surgeons and the public to compare devices, it has the potential to contain insufficient data on clinical outcomes. Therefore, failing implants

may not become apparent until after the three-year mark. This is highlighted by the recent failure of DePuy's ASR resurfacing device which was awarded an ODEP rating of 3A[46] prior to failure rates increasing, being identified as an outlier, and subsequently withdrawn from the market[70]. Despite the potential excellence of the ODEP system, it is imperative that newly introduced implants can be effectively monitored for the early identification of failing devices to reduce significant risk and morbidity to patients.

The Balliol Collaboration is a network of methodologists and clinicians who held a series of conferences at Balliol College, Oxford between 2007 and 2009[110,111]. Their work developed into the IDEAL collaboration[112] designing a framework for surgical innovation. This framework involves a phased introduction of new surgical devices following a 5-step process: Idea, Development, Exploration, Assessment and Long-term follow-up[113] (Table 1-1).

Table 1-1: The 5-step process for the phased introduction of new surgical devices.

Phase 1 Idea	**Phase 2a Development**	**Phase 2b Exploration**	**Phase 3 Assessment**	**Phase 4 Long-Term Monitoring**
Proof of concept	Safety	Efficacy	Comparative effectiveness	Quality assurance
Case reports or first-man studies	Prospective development studies	Prospective observational studies or feasibility RCTs	RCTs	Observational studies in comprehensive registries

These 'IDEAL Recommendations' were developed to acknowledge the pathway for surgical innovation is inevitably different to pharmacological developments. Thus there is a need for a "different approach to evaluation"[112]. The importance of using prospective registries is highlighted and recognises the difficulties in performing Randomised Control Trials (RCT). Therefore, IDEAL suggests in the early phases of a new device there needs to be measures in place to evaluate learning curves. Further into the phased introduction, IDEAL recommend an assessment of the comparative effectiveness of the new implant to current standards using RCTs.

There have been calls for the use of surrogate measures to compare new implants to conventional devices that are proven. An example of such a measure is Radiostereometric Analysis (RSA) that was particularly highlighted[105], which is discussed in more detail in Chapter 2. Nelissen et al compared new implants investigated with RSA versus those that were not[105]. This study demonstrated implants investigated using RSA had marked reductions in revision rate of 22-35% when compared to implants with no radiographic assessment. The authors agreed with the IDEAL recommendations that a "phased clinical introduction of new prostheses with two-year RSA results as a qualitative tool could lead to better patient care"[105].

One of the key recommendations highlighted in the phased introduction by the IDEAL statement is the use of registries to monitor new implants long-term. The NJR is a potentially excellent resource and can be linked to both the Hospital Episodes Statistics (HES) and Patient Reported Outcome Measures (PROMs) databases[114]. Such a combined database has the potential for post-market surveillance of newly introduced

implants. In addition, new surrogate tools that can predict long-term outcome to identify failing implants early, can be investigated. To increase analytic potential there are plans to combine international registries including the Australian and Norwegian registries with the NJR to create a larger database[100]. As new devices are invariably implanted in small numbers initially, a combined dataset would perhaps increase the chances of early detection of failing implants. New device regulations have highlighted the need for a Unique Device Identification (UDI) of all implants[90], which could potentially help in establishing an international registry and improve traceability of devices. DePuy have been piloting such a method in the wake of ASR.

Despite the numerous calls for change in device regulation, progression has been slow. It is only now that new regulations are coming into force, a decade after the failings of MoM implants were first highlighted[90]. The reasons for this are unclear but may relate to the rapid expansion of the number of devices introduced into the UK market. In 1996, there were only 62 primary hip replacement implants available for use in the UK[115]. However, this number has substantially increased to 261 by 2011[103]. Furthermore, surgeons can choose various combinations of femoral stem and acetabular components with over 2,800 different combinations currently available[102]. The impact of proposed changes to EU device regulation and the voluntary Beyond Compliance services once implemented, will not be immediately apparent. In the meantime, there have been changes already introduced. While ODEP have continued to use revision as an endpoint measure of outcome, the NJR have started to incorporate patient reported outcome measures of function in the hope of potentially identifying failing implants more quickly[116]. It is imperative to continue using established test

centres for new devices to be introduced and investigated[116]. It was independent studies that first identified MoM failures[62,65].

1.6. Evidence-base for hip arthroplasty devices

A recent systematic review has demonstrated nearly eight percent of all primary hip replacement operations recorded by the NJR in 2011, used a device which had no readily available evidence relating to its safety or effectiveness to support its use in patients. Unfortunately, this figure is likely to be an underestimation of the true problem. The study recognised the majority of evidence that does exist for the other unrated prostheses is of low quality or relates to short-term outcomes only[117]. This is greatly concerning when compared to the widespread publicity surrounding recent safety issues with MoM devices highlighting issues with device regulation[70]. This finding is not unique to the UK. A comparable study of device evidence to this systematic review was conducted on the Catalan Arthroplasty Registry (RACat)[118]. This demonstrated over 9% of femoral stems and 13% of acetabular cups available in the market had no clinical evidence of effectiveness.

Furthermore, it is concerning to note approximately 45% of brands available to be used in the UK for primary hip replacement have no ODEP rating at all[117]. In accordance with ODEP guidelines, such unrated devices should only be implanted as part of an ongoing trial and not routinely used in the NHS. Despite the large proportion of unrated brands available for use, the prostheses used most widely in the UK were implants which are highly rated by the ODEP system[117]. This implies that most Orthopaedic surgeons are using a prosthesis, for primary hip replacement, that is

supported by a good evidence base. The reasons why some surgeons would choose to use prostheses with no available evidence to support their use is not clear. However, some potential reasons include the introduction of new prostheses by manufacturers who have well established and trusted devices already available on the market. Subsequently, the new implants are considered, perhaps somewhat erroneously, to be of a similar high standard as the established device regarding longevity and clinical outcome. An alternative explanation may be a manufacturer introduces a new device that has an 'improved' design feature to an existing well-established implant by way of a minor modification. Unfortunately, history has recorded several examples of such minor design changes leading to disastrous effects on device longevity and complications for patients[94,95].

The evidence base for total hip replacement implants has been investigated previously by Murray, Carr and Bulstrode in 1995[115]. Their study demonstrated only 30% of implants available for use in the UK had evidence to support their use in patients[115]. At the time the authors concluded that there is considerable need for significantly improving evidence collection for devices. They recommended regulation of all medical devices be improved in both the introduction and post-market surveillance stages. Despite their calls for change and the recommendations of other groups[83,101,113], the evidence base for hip replacement has only marginally improved to 21% in the last 18 years[117]. The reasons for lack of significant improvement are not clear but could be related to the rapid expansion in the number of available devices as discussed earlier. Furthermore, the quality of studies within the Orthopaedic literature is generally of low standard. The systematic review investigating device evidence base identified only

six randomised controlled trials which represents a vastly small proportion compared to the 261 implants available at the time[117]. There are challenges to initiate and coordinate well conducted RCTs for Orthopaedic implants compared to the pharmaceutical industry or for other interventions. This could contribute to the paucity of high-quality evidence for hip replacement implants.

The majority of concerns raised here relate to the process in bringing a new device to the market and the evidence required for this to happen. At the time of MoM failings, there was no requirement for a list of currently available medical implants or devices, for pre-market data, or peer-reviewed publications to be made available publicly. However, new changes to regulations due to come into force have introduced the requirement for this to be made public. It is hoped that such transparency in the early stages of device introduction could allow surgeons and patients to make a more informed decision on which implant to choose.

1.7. Conclusion

Device regulation has come under the spotlight since the public failings of MoM devices and new proposals have been made which should come into force soon. Despite national registries that are excellent resources to identify failing implants, there are several issues of voluntary reporting and lack of ability for early identification. Despite the methodology for comparing hip implants being superior to other areas of device regulation, there remains uncertainty behind the evidence of clinical effectiveness of devices that are available on the market for surgeons. Only 49% of devices implanted overall in 2011 achieved an ODEP rating of 10A/B/C. NICE has a clear benchmark for hip replacement implants of an ODEP rating of 10C, representing the minimal clinical standard for general implantation in the UK[119]. Nearly one in four prosthesis brands available in 2011 for surgeons lacked evidence to support their use in patients. Therefore, there is a need to improve current device regulation with a more rigorous approach, to avoid perpetuating the widespread and uncontrolled manner of a significant number of devices with no available evidence being used. The use of surrogate markers to identify failing devices early is one potential solution, however, it is unclear which are available and validated.

1.8. Scope of thesis

This thesis focuses on the need for and the investigation of novel surrogate markers of outcome in primary THA surgery to be used as part of post-market surveillance to aid the early identification of failing devices. The thesis begins with a systematic review investigating the currently available validated surrogate markers of primary THA devices against long-term revision rate. The thesis continues with the investigation of novel analgesic and healthcare utilisation surrogate markers with the analysis of a retrospective cohort linking data between joint registry and primary care databases. The thesis ends with further investigation of the differences in rates of revision and surrogate marker utilisation between metal-on-metal and other bearings.

1.9. Aim of thesis

- Chapter 2 aims to investigate which available surrogate markers have been validated against the current NICE guidelines for an established THA implant.
- Chapter 3 provides the main methodology and literature review for investigating the novel surrogate markers of primary THA surgery. It aims to investigate analgesic and healthcare utilisation within the first post-operative year following THA as surrogate markers of implant survival.
- Chapters 4 aims to investigate differences between metal-on-metal and other bearings against revision rate and the novel surrogate markers of primary THA surgery.

2. Systematic Review of Current Surrogate Markers of Long-Term Outcome in Primary Total Hip Arthroplasty[§]

2.1. Summary

The public failure of Metal-on-Metal devices have reiterated the need for robust systems to identify failing devices newly introduced to the market. Current systems to analyse failure rates are inadequate to detect failing devices early. This study investigates available validated surrogate markers of long-term outcome in primary Total Hip Arthroplasty. A systematic review was conducted evaluating surrogate markers of ten-year revision rate as the gold standard defined by the National Institute of Health and Care Excellence guidelines. Database searches included Medline, Embase (OVID), Cochrane and Google scholar. Articles were included from the date of their inception to the 8th of June 2015. Initial search yielded 1082 studies of which 115 were included for full review. Only 17 demonstrated surrogate markers validated against long-term outcome including one systematic review, one randomised control trial, one case-control study, and 13 case series. The validated surrogate markers identified measured device migration and wear using either Radiostereometric Analysis (RSA) or einzel-Bild-Röntgen-Analyse (eBRA). This included six RSA and four eBRA studies. Patient Reported Outcome Measures (PROMs) was noted as a potential surrogate marker but not validated against long-term outcome. This systematic review identified two validated surrogate markers of ten-year revision rate in primary hip arthroplasty: RSA and eBRA. It is recommended such validated markers are used in pre-market testing for all new devices. Further studies are required

to investigate PROMs against long-term outcome for its potential use in post-market surveillance.

§Work described in this chapter has been published as[120]:

Malak T T, Broomfield JAJ, Palmer AJR, Hopewell S, Carr A, Brown C, Prieto-Alhambra, Glyn-Jones S. surrogate markers of long-term outcome in primary total hip arthroplasty: A Systematic Review. British Joint Research (BJR) March 2016. Re-use permitted by BJR. The text includes minor and major additions, and formatting changes to the original text. Figures or Tables may have been reproduced with edits.

2.2. Introduction

The public failure of Metal-on-Metal (MoM) devices have reiterated the need for robust systems to identify failing devices newly introduced to the market early[46]. Regulation of devices has been under scrutiny with a recent systematic review revealing 24% of all hip devices available for use in primary Total Hip Arthroplasty (THA) in 2011 had no evidence to support their clinical effectiveness[117]. Of note, this study was unable to explore any clinical evidence that may have been submitted by the manufacturer for market approval. There is even one example of a fictitious company able to gain approval for a fake device claiming similarity to three other controversial devices (two were recalled and one subject to legal action at the time of the publication)[121]. In light of these, recent changes have been implemented in the hope of improving safety[90]. However, there have been calls to improve post-market surveillance of hip devices in addition to changes in pre-market regulation[122].

Current regulatory bodies monitoring new implants, such as the Orthopaedic Data Evaluation Panel (ODEP) and the National Joint Registry (NJR), have been unable to detect failing devices early enough. As part of a unified response to this crisis both the IDEAL group and Beyond Compliance have suggested a phased introduction of new implants with recommendations for post-market surveillance[101,113,123]. However, both schemes are not mandatory. In addition, they rely on collecting basic post-market surveillance data along with revision rates in a similar manner to both ODEP and NJR. Focusing on revision as an outcome has several benefits however, this is a rare outcome and can take several years to occur and thus, registries are potentially best suited for long-term monitoring[122]. New implants are typically introduced in small

numbers and often undergo incremental changes during the early phases[124]. Using revision rate as the sole outcome measure has left regulatory bodies vulnerable in their ability to detect failing implants early[46]. The minimum sample size for a non-inferiority study using contemporary benchmarking standards is 3200, but drops to 1600 with the Kaplan Meier method[125]. However, these figures are markedly higher compared to the 150 participants required as the minimum number for the total cohort to obtain an ODEP rating of 3A*, or the 250 participants for a rating of 5A*[126]. Furthermore, ODEP had given the Articular Surface Replacement (ASR, DePuy, Raynham, Massachusetts) implant a 3A rating, which is the highest rating for its stage of monitoring, despite subsequently being recalled[70].

Therefore, improvements in post-market surveillance of hip devices are urgently sought. The new European regulations require manufacturers to formulate Post-Market Surveillance (PMS) plans for devices[90]. Despite this, there is no specific guidance on the contents or how such a plan would be implemented[127]. PMS presents as one of the few practical methods of detecting failing devices when complications occur infrequently, after long-term use, or are unlikely to be observed until the product has been used in patients[128]. In particular, late implant failures as seen in Orthopaedic devices, are challenging to detect in clinical studies[128]. Furthermore, with incremental changes in the early phase of market introduction, this may be difficult to assess in clinical testing but would possibly be observed during surveillance[128]. These factors associated with hip implants, are making PMS imperative and perhaps more important than pre-market testing when analysing the safety of a new device. Any reform to current post-market surveillance can only be delivered with the aid of a cost-effective,

simple surrogate outcome measure that can yield clinically useful predictions of future function within one to two years of an implant being introduced.

The aim of this systematic review is therefore to investigate which available surrogate markers have been validated against the current National Institute for Health and Care Excellence (NICE) guidelines for an established THA implant.

2.3. Methods

2.3.1. Objective

A systematic review was performed to investigate whether any validated surrogate markers of long-term revision for primary THA are currently available[129].

2.3.2. Eligibility criteria

A systematic review of the literature was conducted following the Cochrane guidelines. Eligible studies investigating a surrogate marker in predicting long-term outcome in primary THA *in vivo* by systematic review, Randomised Control Trials (RCT), cohort analysis, or case-control analyses were eligible in this study.

Long-term outcome was defined as revision rate of a device at ten years in accordance with NICE guidance, which considers a proven implant is one with a revision rate of

less than 5% at ten years[130]. In addition, this systematic review included studies that compared potential surrogate makers against subsequent revision rate from seven to ten years. All modes of revision were investigated excluding fracture, dislocation and infection as other factors are likely to contribute to such failures rather than solely being implant related[131,132].

This systematic review defined a surrogate marker as a measure that can reasonably predict a clinical outcome of interest, irrespective of whether it directly measures that clinical outcome[133]. A surrogate marker is different to a risk factor of an outcome. A risk factor or predictor is present prior to surgery and is associated with the event of interest (revision in the case of THA). In contrast, a surrogate marker is present after the operation and is associated with the event of interest. Therefore, this systematic review specifically investigated surrogate markers, measured post-operatively, that have been shown to be associated with long-term outcome revision rate.

Studies that used inferential statistics (e.g., regression, chi-squared, sensitivity and specificity, ROC curve univariable or multivariable analysis) to investigate the measure of a surrogate marker compared with long-term revision rate were included in the systematic review.

Studies that involved fracture, dislocation, infection, revision THA, animals, *in vitro* investigations, or used an outcome other than revision were excluded.

A validated surrogate marker was defined as any measure within the first two post-operative years that was statistically shown to be associated with overall implant revision rate at ten years.

2.3.3. Search strategy

A protocol was not published, nor was this systematic review pre-registered prior to its commencement. For the Medline and Embase (OVID) databases a broad search strategy was developed. For the Cochrane database and Google Scholar separate search strategies were adopted. The search strategy terms for Medline, Embase and Cochrane databases can be found in Table 2-1.

Table 2-1: Search strategies employed for Medline, Embase and Cochrane databases. Edited with permission from BJR.

MEDLINE	EMBASE	COCHRANE
1. exp ARTHROPLASTY, REPLACEMENT, HIP/; 17320 results.	20. exp TOTAL HIP ARTHROPLASTY/ OR exp TOTAL HIP PROSTHESIS/ OR exp TOTAL HIP REPLACEMENT/; 22279 results.	1 MeSH descriptor: [Arthroplasty, Replacement, Hip] explode all trees 1616
2. (hip* ADJ3 replacement*).ti,ab; 11179 results.	21. hip* ADJ3 arthroplast*).ti,ab; 18277 results.	2. outcomes:ti,ab,kw (Word variations have been searched) 201113
3. (total ADJ3 hip ADJ3 replacement).ti,ab; 7138 results.	22. 20 OR 21; 30071 results.	3 MeSH descriptor: [Treatment Outcome] 1 tree(s) exploded 95982
4. (total ADJ3 hip* ADJ3 arthroplast*).ti,ab; 12131 results.	23. exp TREATMENT OUTCOME/ OR exp OUTCOMES RESEARCH/; 1017012 results.	4 joint revision 651
5. (hip* ADJ3 arthroplast*).ti,ab; 16151 results.	24. outcome*.ti,ab; 1309132 results.	5 revision rate 1722
6. 1 OR 2 OR 3 OR 4 OR 5; 29953 results.	25. "patient outcome*".ti,ab; 36798 results.	6 MeSH descriptor: [Prosthesis Failure] this term only 589

7. exp TREATMENT OUTCOME/ OR exp TREATMENT FAILURE/; 673333 results.	26. exp PROGNOSIS/; 469482 results.	7 prognosis 18172
8. outcome*.ti,ab; 964585 results.	27. exp PROSTHESIS FAILURE/; 27986 results.	8 MeSH descriptor: [Reoperation] this term only 1647
9. exp PROGNOSIS/; 1122139 results.	28. exp REOPERATION/; 52851 results.	9 MeSH descriptor: [Patient Outcome Assessment] this term only 13
10. exp PROSTHESIS FAILURE/; 21934 results.	29. (joint* ADJ3 revision).ti,ab; 362 results.	10 #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 214295
11. exp REOPERATION/; 67049 results.	30. (revision ADJ3 rate*).ti,ab; 2631 results.	11 MeSH descriptor: [Forecasting] this term only 505
12. (joint* ADJ3 revision).ti,ab; 319 results.	31. 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30; 2297641 results.	12 prediction 5437
13. (revision ADJ3 rate*).ti,ab; 2220 results.	32. exp PREDICTION/; 230887 results.	13 predictor 7330
14. 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13; 1828944 results.	33. prediction*.ti,ab; 228449 results.	14 predict* ADJ3 outcome* 833
15. prediction*.ti,ab; 196780 results.	34. predictor*.ti,ab; 311190 results.	15 #11 or #12 or #13 or #14 13100
16. predictor*.ti,ab; 223411 results.	35. (predict* ADJ3 outcome*).ti,ab; 72916 results.	16 #1 and #10 and #15 3
17. (predict* ADJ3 outcome*).ti,ab; 50942 results.	36. 32 OR 33 OR 34 OR 35; 647538 results.	
18. 15 OR 16 OR 17; 425222 results.	37. 22 AND 31 AND 36; 578 results.	
19. 6 AND 14 AND 18; 467 results.		

For Google Scholar database the following search terms were used: prediction / long-term / ten-year / outcomes / hip arthroplasty / hip replacement. Searches were performed to include articles from the date of their inception to the 8th of June 2015 for each of the databases. Despite no restriction of language to the inclusion criteria, any article for which a translation was unavailable was excluded from this systematic review. No articles are excluded for this reason.

Initial screening of articles identified from the search strategy, based on the title and abstract, was independently performed by two reviewers (TTM, JAJB). Following this, full articles were obtained for review. These were examined according to the inclusion criteria. Further screening was performed on the references of any relevant articles for additional articles to be identified and potentially included. A third author (AJRP) resolved any discrepancies.

2.3.4. Statistical analysis

Reviewers used a standardised data extraction form (Appendix 1) to obtain information concerning each eligible study identified during screening. Information that was extracted included: year of publication, study design, level of evidence, sample size, type of surrogate marker investigated, and primary outcome. Statistical significance was included whenever this was calculated. Data extracted from relevant articles was analysed independently by the two reviewers. Completed data extraction forms were compared separately to minimise errors.

A meta-analysis of included studies was not performed due to their limited number and the heterogeneous nature of these studies. In addition, it was not possible to perform risk of bias assessments as the majority of included studies were not trials. There were two systematic reviews included for which a risk of bias was performed within their respective studies.

Themes of validated surrogate markers of long-term outcome in primary THA were identified. These findings were reported descriptively. A statistical value of significance of $p \leq 0.05$ was accepted for determining the predictive validity of a surrogate marker and establishing an association with revision rate.

2.4. Results

2.4.1. Outcome of search strategy

This systematic review performed initial searches on Medline, Embase, and Cochrane databases identifying 1048 potentially eligible studies. Searches of the references of these studies and of Google Scholar identified a further 35 studies. Thus, following initial searches a total of 1083 potentially eligible studies were identified. No further relevant trials or studies according to the inclusion criteria were identified after searches in the ClinicalTrials.gov and NHS evidence databases. Removing duplicate articles and application of the exclusion criteria, as outlined above using title and abstract, excluded 967 studies leaving 116 remaining for full article assessment. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) details the screening and exclusion and can be found in Figure 2-1.

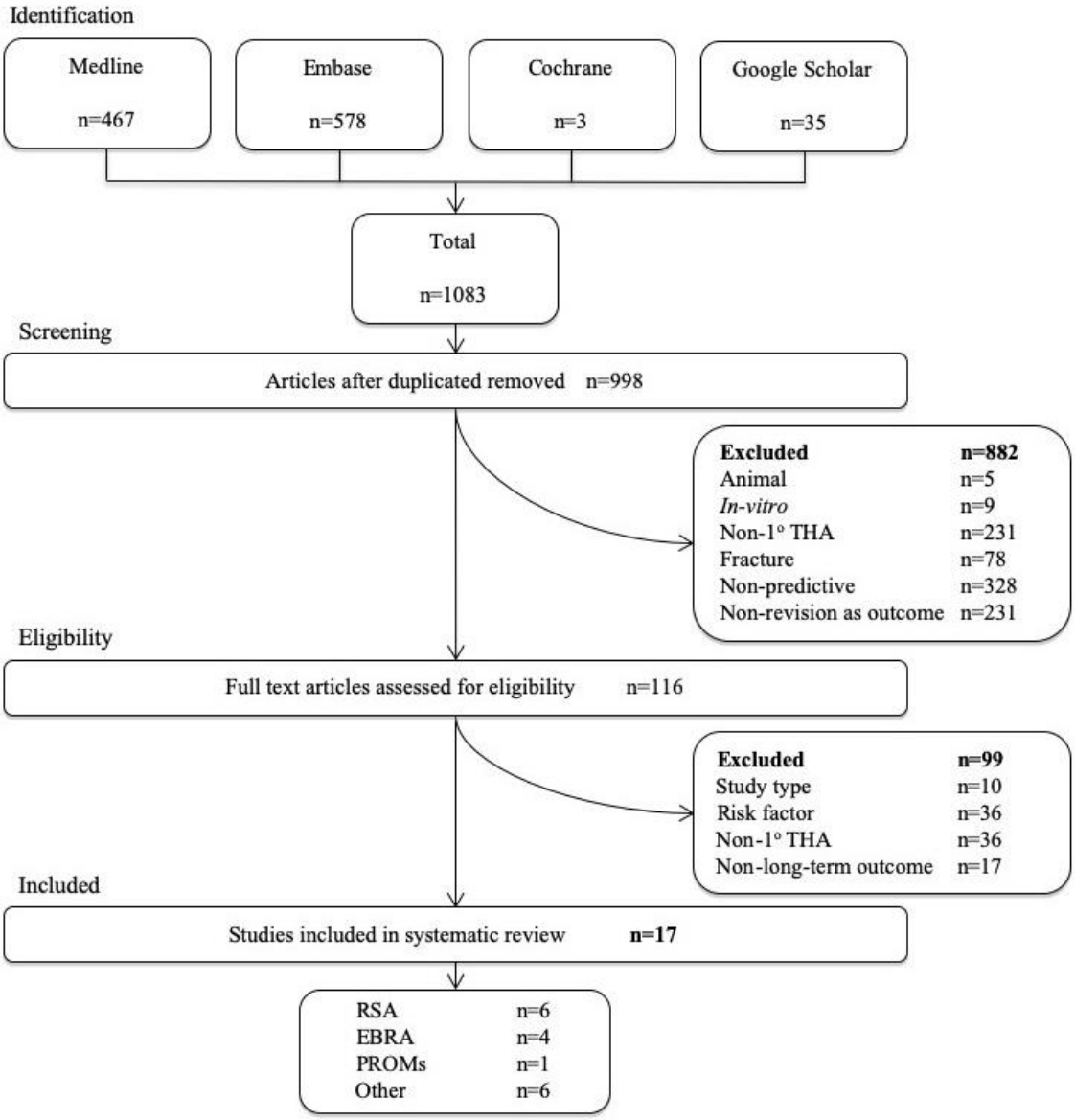


Figure 2-1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow-chart. Edited with permission from BJR.

Following review of full articles, a total of 17 studies were identified that met the inclusion criteria for this systematic review investigating the validity of a surrogate maker against long-term outcome in primary THA (Table 2-2). Studies included were two systematic reviews, one Randomised Control Trial (RCT), one case control and 13 case series.

Table 2-2: Characteristics of included studies. Edited with permission from BJR.

Reference	Design	Sample Size	Surrogate Marker	Implant Type	Follow-up	Primary Outcome	Correlation to Revision	Result
Sesselmann S et al 2013[134]	Systematic Review	309	RSA	All	10	Revision Rate *vs* 2-year Migration	Y	None
Pijls B G[135]	Systematic Review	700	RSA	Acetabular components	10	Revision Rate *vs* 2-year Migration	Y	Migration of 1.0 mm was considered unacceptable
Hauptfleisch J et al 2006[94]	Case Series	118	RSA	Cemented	9	Revision Rate *vs* 2-year Migration	Y	Significant difference in posterior head migration p = 0.002
Nieuwenhuijse M et al 2012[150]	Case Series	39	RSA	Cemented	10	Revision Rate *vs* 2-year Migration	Y	Continued subsidence associated with higher revision p < 0.001
Karrholm J et al 1994[136]	Case Series	84	RSA	Cemented	7	Revision Rate *vs* 2-year Migration	Y	Subsidence > 1.2 mm at two years associated with higher revision p < 0.000005
Freeman M et al 1994[137]	Case Series	206	RSA	Cemented	7	Revision Rate *vs* 2-year Migration	Y	Subsidence > 1.2 mm at two years associated with higher revision p < 0.003
Krismer et al 1999[141]	Case Series	240	EBRA - Femoral	Uncemented	10	Revision Rate *vs* 2-year Migration	N	Subsidence > 1.0 mm at two years associated with higher revision p = 0.0001
Krismer et al 1996[139]	RCT	120	EBRA - Acetabular	Uncemented	8	Revision Rate *vs* 2-year Migration	Y	Subsidence > 1.5 mm at two years associated with higher revision p = 0.005
Mazoochian F et al 2007[142]	Case Series	10	EBRA	Uncemented	7	Revision Rate *vs* 2-year Migration	N	None (Descriptive)

Hendrich C et al 2006[140]	**Case Series**	**11**	**EBRA**	**Uncemented**	**10**	**Revision Rate *vs* 2-year Migration**	**Y**	**Subsidence >1.0mm at two years associated with higher revision p = 0.002**
Takenaga R et al 2012[143]	Case Series	50	PROMs Functional Tests	Uncemented	10	Revision Rate *vs* 21 Day Test	N	No association p>0.05
Kobayashi A et al 1997[144]	Case Series	575	XR Migration	Plain Radiograph Assessment	10	Revision Rate *vs* 2 Year Assessment	Y	Subsidence > 2.0 mm at two years associated with higher revision p < 0.0001
Ranawat C et al 1995[145]	Case Series	236	XR Bone-Cement Interface	Plain Radiograph Assessment	10	Revision Rate *vs* Bone-Cement Interface	Y	Observer rated radiographs deemed as loose at 1 year associated with higher revision rate p = 0.002
Dowd J et al 1994[146]	Case Series	48	Digitizer Programme - Migration	Uncemented	10	Revision Rate *vs* Digitizer Programme	Y	Subsidence > 0.3 mm per year p < 0.001
Malik M et al 2005[147]	Case Control	201	Post Op XR (Wear in Gruen Zones)	Cemented Uncemented	10	Revision Rate *vs* Post Op XR	Y	Thickness of cement mantle in Gruen zones 6 and 7 associated with higher revision rate p = 0.040 and p = 0.003 Radiolucent lines in zone 3 and 5 p = 0.0001
Scott G et al 2006[148]	Case Series	143	XR Migration using Digitizer	Cemented Uncemented	10	Revision Rate *vs* XR Migration at 3 years	Y	Subsidence >3.0mm at three years associated with higher revision p < 0.0001
Khalily C et al 1998[149]	Case Series	119	XR Assessment (Wear in Gruen Zones)	Uncemented	10	Revision Rate *vs* 2-year XR Assessment	Y	Wear in Gruen zones 1,7,8 and 14 at 2 years associated with higher revision rate p < 0.001

The surrogate markers identified in this systematic review are implant migration and wear, and patient reported outcome measures (Table 2-3).

Table 2-3: Validated surrogate markers of long-term revision rate in primary total hip arthroplasty.

Surrogate marker	N	Predicts revision rate	Does not predict revision rate
Radiostereometric Analysis – Migration and Wear (RSA)	6	6[94,134–138]	-
Einzel-Bild-Röntgen-Analyse – Migration and Wear (eBRA)	4	2[139,140]	2[141,142]
Functional Outcome Score	1	-	1[143]
Wear on standard radiographs	6	6[144–149]	-
Total	**17**	**14**	**3**

2.4.2. Implant migration and wear

Implant migration and wear of both the acetabular and femoral components was investigated in ten of the 17 eligible studies. These validated the use of migration measured within the first two years post-operatively as a surrogate marker for subsequent failure (Table 2-2 and Table 2-3). There are two main techniques identified in this systematic review that are commonly used to measure migration: Radiostereometric Analysis (RSA) and einzel-Bild-Röntgen-Analyse (eBRA).

RSA was investigated in six of the eligible studies and found to predict long-term revision rate (two systematic reviews[134,135] and four cases series[94,136,137,150]). Femoral stem migration more than 1 mm to 2 mm within two years post-operation was associated with an increased rate of revision in two studies[94,134]. Cementless femoral stems are

designed to have minimal or no migration, therefore, any further migration is associated with increased risk of higher revision rates[134]. In contrast, cemented femoral stems are expected to undergo initial migration within the first two post-operative years. However, more than 2 mm in two years correlates with a higher revision rate[134]. It was noted that particular migration was significant for different types of fixations. For cementless stems, distal migration correlated with revision rate, whereas for cemented components posterior migration of the femoral head was associated with failure. Migration of the acetabular component was also measured ranging from 0.17 mm to 0.24 mm in the first post-operative year. Migration of more than 0.1 mm per year was deemed significant in all studies investigating wear of the acetabular component.

Mean accuracy for RSA in clinical studies was investigated in one systematic review. The mean accuracy was found to be between 0.1 mm and 0.2 mm for translation migration with a range from 0.005 mm to 0.5 mm. Rotational migration had a mean accuracy between 0.15 degrees to 1.15 degrees[134].

RSA was not the sole method of measuring migration with four studies investigating the use of einzel-Bild-Röntgen-Analyse (eBRA) to analyse migration of both the acetabular and femoral components (one RCT[139] and three case series[140–142]). All studies reported accuracy of migration. eBRA had an accuracy of 1.0 mm to 1.5 mm in measuring the migration of the femoral component. For the acetabular component, the mean accuracy was 0.8 mm to 1.0 mm. One study compared the accuracy of eBRA to

RSA with a mean difference of 0.08 mm when measuring the acetabular component for wear[139].

2.4.3. Other surrogate outcome markers identified

There were six studies meeting the eligibility criteria which identified other surrogate markers of long-term THA outcome (one case control[147] and five case series[144–146,148,149]). Again, these methods involved investigating wear but using standard radiographs. These methods were associated with subsequent revision rate. However, both RSA and eBRA have largely superseded these methods due to their accuracy and precision.

2.4.4. Functional outcome score

In one study, functional outcome scores were investigated for any association with long-term revision rate (one case series[143]). Functional outcome scores included the Harris Hip, SF-36, 6-minute walk and Tegner Lysholm scores. However, there was no significant association found with subsequent revision rate at ten years[151].

2.5. Discussion

This study systematically reviewed the literature for potential surrogate markers investigated for their association with long-term outcome in THA. The primary measure of surrogate marker found to be validated against subsequent revision rate was implant migration and wear within 2-years post-surgery. The use of imaging techniques to measure wear and migration seems to have the most promise for identifying failing devices early. A high degree of accuracy is required in each

dimension of at least 0.1 mm to 0.2 mm[152,153]. There were two main validated methods for measuring migration and wear measurements: RSA and eBRA.

Acetabular wear measurements can be measured by eBRA which is a two-dimensional imaging technique. Programmes are available to measure migration and wear for both the acetabular and femoral components[154]. This study found mixed evidence to support eBRA measured migration and wear and subsequent revision rate with two studies showing correlation and two which did not[139–142]. Furthermore, eBRA was found to be accurate when measuring wear for acetabular components but accuracy was poor for femoral migration. Adequate accuracy to measure acetabular component migration and wear was shown with a threshold 0.1 mm per year[155]. However, accuracy in measuring migration of the femoral component was only 1 mm to 1.5 mm[134]. Therefore, its use as a surrogate marker of subsequent revision for all new THA devices may not be adequate especially at identifying failing implants due to femoral wear. In contrast, RSA had excellent accuracy and precision in measuring both the acetabular and femoral components for both wear and migration. Therefore, using imaging methods as a surrogate for subsequent revision rate has shown the most promise with RSA being the most accurate.

RSA follows on from the basic principles developed in 1898 by the radiologist Davidson[156] in determining the relative movement of two distinct objects using two-dimensional imaging. However, such techniques are complex. RSA is a three-dimensional imaging technique which uses a cage to construct a matrix for which tantalum markers that are inserted in the femoral and acetabular bone can be aligned with special coordinates according to that matrix. Thus, relative movement of the

components against the bone can be calculated to a high degree of accuracy[157]. Some methods required tantalum markers on the implant but newer methods such as that used in Oxford[158], are able to identify the implant digitally without the use of such beads.

RSA is an attractive solution for monitoring and screening new hip devices. There have been recommendations for RSA to be used as part of the introduction for all new implants[159]. There are others who consider RSA as the gold standard surrogate marker for predicting long-term outcome in hip arthroplasty[105]. Due to the high accuracy, relatively few patients are needed to conduct randomised controlled trials for meaningful conclusions to be drawn about later revision rate[134,160]. Furthermore, these conclusions can be ascertained early within two years of commencing the trial[161]. Unlike eBRA, RSA is able to monitor migration and wear accurately for all fixation types of both the acetabular and femoral components including the wear of hard-on-soft bearing surfaces[158]. In addition, RSA can detect failing implants due to aseptic loosening that are otherwise not identified by any joint registry which are unable to detect such devices unless they are revised.

Despite the clear benefits and the potential to be used as a screening tool for all hip implants newly introduced into the market, RSA has significant limitations. RSA requires specialist expertise using bespoke equipment and is thus difficult to implement. Therefore, RSA is only performed in a few centres around the world limiting the ability for it to be used as a screening tool[162,163]. Furthermore, some RSA techniques require marker beads to be inserted onto the hip device to allow for

detection within the cage system[162]. This could impact implant design which ultimately may differ to that introduced into the market and subsequently used in patients on a larger scale.

There are no universally accepted thresholds for RSA migration as each system is different, nor are there time frames for such thresholds established[164]. Therefore, there is significant reliance on expert interpretation limiting where RSA can be perfomed[165]. Thus, RSA is further limited to few research centres with these expertise[166]. New implants that are subject to RSA testing can be compared to a proven, conventional implant but no thresholds for what constitutes a significant clinical difference exists. In addition, there are no universally accepted criteria for tolerable migration[164]. The majority of studies do not follow the recommendations presented out by Kärrholm[167].

While small sample size is advantageous with regard to potential harmful impact to patients and costs to manufacturers, there have been concerns. Firstly, predictive power of the study can be diminished[164]. Secondly, generalisability may not be possible in such a small sample size[168]. Thirdly, a small sample can be heavily influenced by outliers and can skew results[169]. Therefore, it is difficult for RSA to truly be predictive of long-term outcome for a new implant representing its use in a larger population.

One of the most significant limitations for RSA to be used as a potential screening tool is that it has failed to "universally predict long-term" fixation following a recent study[164]. In this study, RSA was unable to correctly predict subsequent revision rate

and stability in two out of eleven cases. For one of the implants RSA predicted long-term stability where subsequent long-term revision rates were above the threshold for acceptability. For one device, RSA predicted it would be unstable but subsequent revision rates were acceptable. The Furlong stem had shown higher long-term revision rates resulting in the Australian Joint Arthroplasty Registry (AJAR) recommending that is no longer used in 2016[170]. However, RSA showed the Furlong stem to be stable[171]. The AJAR has shown the VerSys stem to have acceptable long-term revision rates and ultimately demonstrated to be non-inferior to proven devices[170]. However, RSA predicted the VerSys to be unstable with migration higher than threshold levels and that it would perform poorly[172].

RSA detects THA failings as a result of aseptic loosening by using migration as surrogate measure and thus may be inadequate to detect other modes of failure[164]. In a study performed by Penny et al, RSA demonstrated the ASR implant as stable[173]. Unfortunately, the mode of failure for metal-on-metal implants was not aseptic loosening and thus may not be detectable by measuring migration or wear[46].

Although this study showed no validation against long-term revision rate, Patient Reported Outcome Measures (PROMs) and functional assessment at day 21 post operatively was shown as a potential alternative to measuring wear or migration. There was no evidence the Harris Hip Score could predict subsequent revision rate, however, the Oxford Hip Score (OHS) had shown promise[174]. In this study, patients with an OHS score of less than 27 at six months post operation was associated with higher early revision rates up to two years[174]. This systematic review had found no studies investigating PROMS as a surrogate marker against long-term revision rates.

PROMS is an attractive potential surrogate marker for hip implant failure and has benefits surpassing RSA. The cost of RSA is high, estimated at approximately £250,000 per trial for every new device being tested. In addition, RSA requires expertise and specialist equipment due to the technical challenges to perform. In contrast, PROMS is routinely collected in the United Kingdom for all patients who have had THA, making it a simpler and therefore, less costly alternative to RSA[175]. However, to the date of this study, only the OHS measured at six months post-operatively has currently been shown to have any association with subsequent revision rate up to two years[174]. Therefore, further research investigating PROMs as a potential surrogate marker of long-term revision rate in THA is warranted due to its cost effectiveness. In a similar manner to RSA, PROMs may not be a universal method to capture all failing implants. PROMs in its current format may lack the ability to detect failing implants early as implant loosening is a gradual process and symptoms may only present in the final stages[176]. A potential barrier to investigating PROMs is the large numbers required for such a study to validate its use as a surrogate marker to ensure there is sufficient power for statistical calculations. This would result in any new implant being investigating requiring thousands of patients which would have significant cost implications and potential safety concerns should such an implant subsequently be recalled. In contrast, RSA only requires a small number of patients to predict long-term revision rate in a relatively small follow-up time of two years. This makes RSA more appropriate for pre-market testing compared to PROMs. Despite this, there may be scope for PROMs being used as a measure in post-market surveillance to monitor implants after their general introduction to the market and approval for widespread use.

2.5.1. Limitations

A strict and narrow inclusion and exclusion criteria was used for the search strategy of this study. Only studies investigating a surrogate marker that predicts long-term outcome, set at ten years post-operatively, were accepted. Those studies that investigated surrogate markers against earlier revision rates, less than ten years post-operatively, were excluded from this study. Therefore, the search strategy may have missed several potential surrogate markers of failure rates in THA that require further investigations. Although not formally included within the search strategy, this study had analysed studies that investigated a surrogate marker against mid-term revision rates, but no additional information was yielded.

In addition, the term 'predict' was used in the search strategy which could have excluded studies investigating potential surrogate markers with a focus on a causal pathway or association with failure. Due to a lack of available indexing terms and a universally accepted definition for surrogate markers, it would still remain possible to miss potential studies. By strict definition a surrogate marker is different to a predictor. The term surrogate marker can be defined as a measure that can reasonably 'predict' the clinical outcome of interest. Therefore, the term predict was used in this search strategy. Preliminary research confirmed that the term 'predict' was used most often to infer a potential surrogate marker of outcome in the majority of studies prior to the application of this study's search strategy.

Revision rate was used as the outcome of interest, but this measure can exclude important implant failures. There are patients who have declined revision surgery, perhaps are unfit to undergo such complex surgery, or may not have sought medical

advice for their symptoms. Therefore, the measure of an implant performance may be skewed. However, the gold standard for a proven implant is its revision rate at ten years and thus this study used this definition as the outcome.

Only studies that were published have been included in this study. Therefore, there is the potential for publication bias as it is possible for only those studies with statistically significant results being published by the authors.

This study was unable to combine results of included studies due to the lack of homogeneity. Furthermore, due to the varied reporting of included studies the methodological quality for each was not analysed. As previously mentioned, the surrogate markers relying on imaging techniques are limited to only detecting migration or wear of an implant which is the pre-dominant method of failure for most THA devices. Alternative or unanticipated modes of failure would not be detectable via measurements of wear and migration. Such modes include soft-tissue reactions causing local damage before loosening is present, such as occurred with metal-on-metal devices. A further limitation of imaging techniques is their accuracy to measure wear in harder bearing surfaces, such as those made from metal or ceramic materials.

2.5.2. Clinical relevance

The failures of MoM devices have been discussed previously as a trigger for new regulations[94,177–179]. As a result there have been calls for a phased introduction of all

new THA devices which should be assessed using techniques such as RSA[180,181]. Although RSA can predict later revision rate for many THA implants, it is not a universal predictor for a device's subsequent performance. RSA can only detect migration or wear which is a surrogate for the mode of failure: aseptic loosening[182]. This is most evident by a recent study investigating the recalled ASR implant where RSA failed to predict failure[183]. This is due to RSA being unable to detect other potential modes of failure such as trunnion corrosion, elevated metal ions, frictional torque, or edge loading[183].

Therefore, due to the limitations of using migration and wear measurements as a surrogate marker of outcome in THA, pre-market investigations of new implants using RSA must be used in conjunction with robust post-market surveillance[182]. While the NJR provides an excellent source of information with full coverage of all THA operations undertaking in the UK, and ODEP requiring clinical evidence for a rating, both were unable to detect the failings of MoM devices early. There have been changes as a result of the MoM fiasco with the ODEP and NICE benchmarks being reduced from ten to five percent revision rate at ten years for a proven implant[108]. Unfortunately, there are concerns with failing devices up to five-years post-market deployment with notable devices receiving a good rating after three years[46]. Therefore, alternative surrogate markers to RSA are urgently required to ensure failing new implants are identified within one or two years of surgery to reduce risk and complications to patients.

This study has implications for the regulation and commissioning of orthopaedic devices. Currently, the most accurate and precise surrogate marker that has been validated against long-term outcome for both acetabular and femoral components of THA is RSA. Unfortunately, RSA is not a universally accurate predictor and can only detect one mode of failure (aseptic loosening). It is unable to detect other modes, a fact that has been recently highlighted by the MoM devices[173]. However, this study still recommends the use of RSA in the evaluation of all new implants prior to their general release to the market place, as part of a phased introduction, in conjunction with other bodies such as NICE and the IDEAL group[123,184]. Furthermore, RSA should be performed on any new implant including new materials or bearing surfaces, irrespective whether they are deemed to be of similar design to an existing device.

2.5.3. Conclusion

In conclusion, this systematic review identified the measurement of wear and migration to be validated surrogate markers of long-term outcome in primary THA. Two methods of measuring wear migration were found: RSA and eBRA. While eBRA was accurate for wear it failed to be validated for migration. RSA was validated for both wear and migration for both the acetabular and femoral components. This study recommends the use of RSA for pre-market testing of all new implants as part of a phased introduction. However, RSA is not a universal predictor of subsequent outcome and therefore must be used in conjunction with post-market surveillance. Further investigation is required for scalable and less resource intensive potential surrogate markers that can be used in post-market surveillance.

3. Investigation of Novel Surrogate Markers of Mid- to Long-Term Outcome in Primary Total Hip Arthroplasty

3.1. Summary

Considering recent THA device failings, the debate on regulation, and lack of cost-effective surrogate markers, the need to investigate alternative surrogate markers is highlighted. To analyse real-world registry data requires complex approaches. This chapter investigates the association between potential surrogate markers using routinely collected data and subsequent revision rate.

3.2. Introduction

Total Hip Arthroplasty (THA) has been hailed as one of the greatest procedures in modern medicine[2]. However, in this thesis the failings of Metal-on-Metal (MoM) implants has been highlighted as one of the examples leading to the recent changes in current device regulations[90]. The regulations present at the time of the MoM failings were discussed in Chapter 1. Since then, new regulations have come into force.

3.2.1. New device regulations

In recent decades there has been a significant increase in the number of medical devices in general[90]. For THA implants there has been more than a fourfold increase in the number of available devices for surgeons to choose from between 1996 and 2011[103,115].

The majority of medical devices are intended to last considerable time *in-situ* with some devices potentially remaining for the remainder of a patient's life[90]. Therefore, the expectation from patients is for these devices to be safe, reliable and perform as they have been designed.

The failings of MoM implants combined with exposure of Poly Implant Prothèse (PIP) implants being fraudulently used led to a call for immediate action by the European Commission[185]. Subsequently two new regulations have come into force in 2017: the Medical Device Regulation (MDR)[186] and the In-Vitro Diagnostic Device Regulation (IVDR)[187]. Rather than redesign the current system, these new regulations serve to revise them to ensure a more modern and robust framework. In the UK, device regulation may differ following departure from the European Union. The new EU regulations are due to fully apply to EU member states in May 2022, after the transition period, and thus will not automatically be applicable to the UK[188]. All devices will need to be registered with the MHRA as a CE mark alone is no longer accepted after 2023[188]. However, the MHRA "has highlighted the importance of strengthened regulations that do more to protect patients" and are likely to adhere to the new EU medical device regulations closely[188].

A significant reason for the inability to identify such failing implants early is perhaps found in the difference between the pharmaceutical and medical device industries regarding the requirement for clinical studies prior to market use. The original requirements for market entry of a device such as used in THA has been discussed in Chapter 1. THA implants are merely required to obtain a Conformité Européenne (CE) mark prior to introducing their device in any market within the European Union[90].

This can be approved from any one of the numerous bodies available[91,92]. However, unlike in the pharmaceutical industry, the standards required were mainly technical in nature with little to no clinical testing[90]. There are certain barriers for medical devices to perform clinical studies that include randomisation, blinding, and the need for long-term follow-up to capture implant failure. Therefore, the validity of clinical studies for medical devices has been called into question which may result in a high-risk of misinterpreting results[189,190]. Following the public disasters of MoM and PIP devices additional features were introduced to the framework to ensure higher quality and safety. Prior to these it was technically possible to introduce a THA implant claiming equivalence to an existing proven implant and therefore be used in patients without clinical evidence of its safety or effectiveness[92].

Thus, one of the main changes in regulation concerns the requirement for clinical evidence prior to the introduction of a new device. These requirements formed part of the 'essential requirements' known as the General Safety and Performance Requirements (GSPR) in the new MDR regulations[90]. The MDR has defined which clinical data to be evaluated by the manufacturer. It will recognise and use all available outcome data and will review data on a yearly basis. Every five years, a process of recertification for the CE mark will be performed. While the notion of 'equivalence' remains in the MDR, the biological, technical, and clinical aspects that need to be demonstrated to qualify for equivalence has been defined. Furthermore, there is now a requirement that if a manufacturer claims equivalence for the new device, they must prove there is no significant difference in safety and performance. As a result, clinical data must be evaluated prior to the implant being issued approval for a CE mark. In

addition, the clinical data can be used to inform the manufacturer of the extent of any further clinical evaluation of the device that is required by either Post-Market Surveillance (PMS) or Post-Market Clinical Follow-up (PMCF) (Figure 3-1).

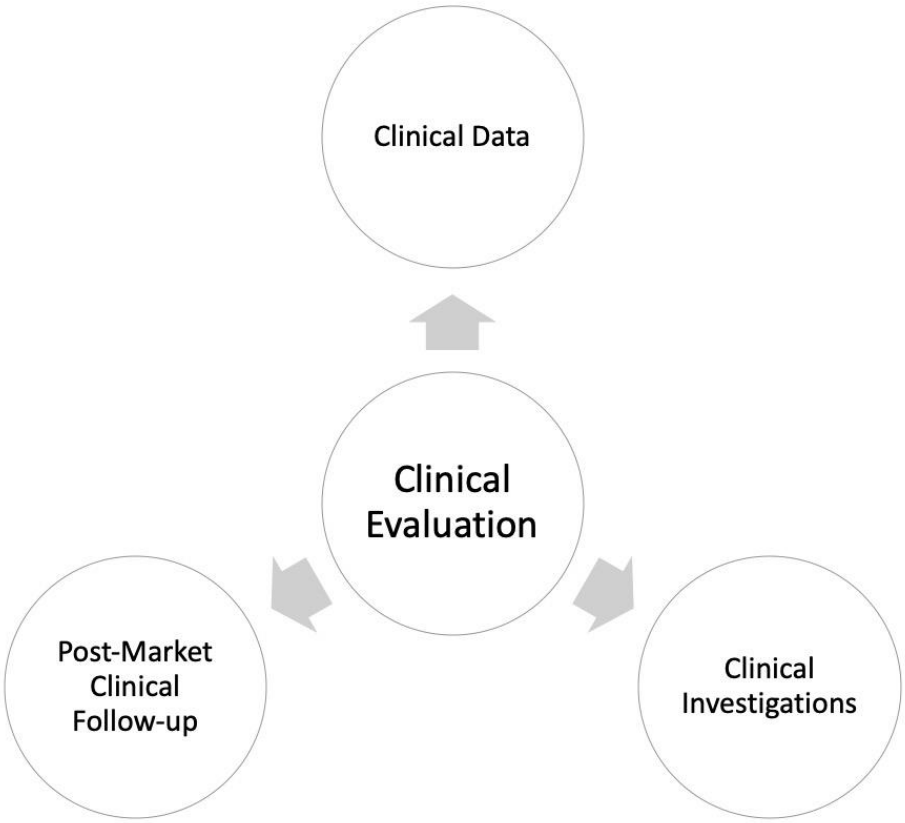


Figure 3-1: Sources of evidence for clinical evaluation of medical devices.

The MDR has repeatedly reported that the clinical data should be used to inform post-market evaluation and "must be sufficient"[90]. To reinforce this, the MDR introduced "Common Specifications" designed to allow for harmonisation of the requirements for clinical evidence, potentially limiting the ability for manufacturers to shop around notified bodies for approval. Furthermore, the MDR introduced the "Clinical Evaluation Consultation Procedure" which has developed the nickname: "scrutiny

procedure"[90]. The scrutiny procedure is mandatory for all devices that are deemed high-risk and includes THA implants. It requires that not only is clinical evidence assessed by the notified body, but it is also evaluated by an expert panel who then have 60 days to provide an opinion in regard to that specific device.

Despite registries used for the surveillance of devices being around for some time, they have never been mentioned in European legislation for medical devices[90]. The MDR has made special note of registries, encouraging all member states in Europe to establish a registry for specific devices including those used in THA. Article 83 calls for the establishment of registries for devices considered high-risk in all members states. It has also provided principles for the collection of data to ensure registries in Europe can collect comparable information. Therefore, potentially unifying the data pool and increasing power of future investigations. In addition, the MDR suggests that these registries be used to assess long-term performance of devices by independent parties.

Interestingly, the MDR goes further than to simply recommend the use of registries. It requires both manufacturers and notified bodies to use data from registries when formulating decisions about continuing use of the device as part of their obligations for post-market monitoring. Manufacturers must participate in registry or PMCF studies with explicit mention for each device they must "plan, establish, document, implement, maintain and update a post-market surveillance system"[90]. Any post-market surveillance system must be aligned with the risk class of the device and be "suited to actively and systematically gathering, recording and analysing relevant data on the quality, performance and safety of a device throughout its entire lifetime, to drawing

the necessary conclusions and to determining, implementing and monitoring any preventive and corrective actions"[90]. Thus, participating in a registry would provide ample opportunity for a manufacturer to meet such requirements. In regard to notified bodies, data from registries is used in part to assess the device during recertification.

Registries offer the possibility to monitor safety and performance using real-world data by systematically collecting data on the effectiveness of devices following market introduction[191]. They provide the opportunity to scrutinise new implants through a naturalistic approach as they are used by surgeons with varying experience, in a variety of institutions, and in patients of varying ages or co-morbidities[192]. The primary outcome measure used in THA implant registries is revision rate used as a marker of performance. One of the benefits to using registries is the ability to compare new devices with similar proven implants. In order to achieve this, data on a wide range of devices needs to be collected and thus registries such as the NJR[104], in which it is mandatory for data on all THA devices to be submitted, provide a valuable resource.

Registries, such as the NJR, were originally designed to help identify patients with failing implants for revision surgery[33]. Prior to this there had been no systematic method to identify patients with such devices for recall despite Sir John Charnley recommending a central registry decades beforehand[102]. The MDR reiterates the importance of being able to urgently recall patients using a registry for traceability of failing devices. There is a call for standardising registries to enable aggregation of data between member states which would require a single method of identifying a device. It is the first time in regulations that a requirement has been made for all devices to have

Unique Device Identification (UDI). UDI may help identify devices in the early stages of introduction to market which can go through multiple iterations. Furthermore, an UDI number will be constructed in a universal manner as specified by coding standards to ensure unambiguous identification across the member states. In the case of THA implants, the MDR requires UDIs are stored by both healthcare institutions and professionals. A registry can help ensure these requirements are met. The MDR ensures greater transparency by requiring the UDI to be included on the patient's implant card in conjunction with other important information relating to the device. This implant card would be provided routinely to the patient.

The MDR goes beyond the recommendations of keeping a registry and has created the European Database on Medical Devices (EUDAMED). The key aspect of the EUDAMED is the integration of the different data collection systems to enhance the analysis of information regarding "economic operators, conformity assessment, notified bodies, certificates, clinical investigations, vigilance and market surveillance"[90]. EUDAMED is divided into three key areas: clinical investigations, vigilance, and post-market surveillance.

There are several advantages of a central database. It can ensure transparency by allowing both healthcare professionals and patients access to current and up to date information regarding a device. Transparency is one of the key aspects highlighted from the failings of devices such as MoM and PIP implants[46]. As a result, the MDR formulated the Summary of Safety and Clinical Performance (SSCP), for devices with a CE mark considered high-risk, which will be accessible on EUDAMED to the public.

These high-risk devices include all class III devices, such as THA implants, any implantable devices, and any device still under investigation. This marks the first-time data used to obtain a CE mark is available to the public. Manufacturers can avoid the need to report to multiple agencies with various bodies such as notified bodies and the member states can have access to a single source of information through EUDAMED. With regard to the vigilance database, manufacturers can report adverse outcomes from multiple markets and member states into a single database allowing for the potential to detect serious issues earlier. Ultimately, the MDR for the first time acknowledges the importance of registries in learning from real world data in real-time. As a result, both the Orthopaedic Data Evaluation Panel (ODEP) and Beyond compliance have collaborated to ensure manufacturers can potentially use registry data that comply with post-market regulatory obligations[90].

3.2.2. Limitations of registries

The MDR has placed an emphasis on post-market surveillance with registries at the forefront of providing a unified early detection system that manufacturers can feed in to from all member states. Furthermore, there has been a focus from the MDR for the need of post-market surveillance. Registries can provide opportunities for long-term surveillance using national, real-world data[193]. In response, there have been calls for collaboration between registries with a notable example being the NJR combining their database with the Endo-Prothesen Register Deutschland (EPRD)[90].

However, registries such as the NJR can be limited in the early detection of failing implants. Long-term follow up is required to determine safety with current guidelines deeming a proven hip is determined by a revision rate at ten years[130]. While registries are arguably suited to be able to collect data required for such monitoring notable failings have occurred revealing its limitations. There are a number of factors contributing to a registry's limitations in detecting failing implants early. These include the inability to monitor adverse events in 'real-time', lack of comparative studies to draw conclusions, minimal data collected for patient characteristics and several others[122,194]. One of the key factors for the failure of registries to detect poorly performing devices early is due to revision rate being the only current measure of long-term monitoring[195]. Revision as an outcome has several advantages. Firstly, it is an excellent marker of quality for a device. Secondly, there are significant health burdens to both patients and society. For patients, revision surgery can result in morbidity such as disability or poor mobility[196,197]. Revision surgery often performs less well than the original surgery[198]. In addition, revision surgery can increase costs for institutions such as the NHS and impact society, for example patients requiring sick leave. Finally, revision is a binary, 'hard endpoint' that is therefore reproducible and comparable making it an excellent outcome measure for studies. Revision data is easily collected compared to other outcome measures and can be collected accurately.

Unfortunately, using revision as an outcome does have significant limitations. Firstly, revisions occur after significant time has lapsed since the operation[199]. In the case of the 3M Capital Hip, 4,000 patients underwent primary THA with this implant in the UK from 1991. Concerns were only raised in 1995[200], and a health hazard warning was

issued in 1998[201], seven years after market introduction. As a result, the indications for revision surgery may not directly correlate as a marker of device quality. There are a number of factors that can influence a surgeon's relative view regarding to the performance of a device. This can include length or quality of recovery for the patient and the ease of performing the operation. Such factors are not recorded within registries nor taken into account when using revision as an outcome of implant performance. In addition, revision surgery is not performed in all patients that may require an operation even for a failed implant due to co-morbidities or patient choice[202]. Surgeons may be unwilling to recommend revision surgery for ageing patients who in turn may be unwilling to proceed with surgery irrespective of the implant outcome quality[195]. Studies have shown no correlation between age and quality of outcome (satisfaction) suggesting a lack of objectivity for deciding to proceed with revision surgery[203]. Furthermore, surgeons may have different thresholds for revision surgery depending on the device. For example, resurfacing or bone conserving THA devices may be regarded as easier to revise and thus more amenable to revision surgery[195]. There are different types of revision which may not be related to the device. Revisions can range from an exchange of the liner to removal of the component[202]. It is not possible to integrate the survivorship between components despite how early revision of one may affect the other, such as removing the femoral component could lead to abnormal forces through the acetabular component. Revision surgery is a relatively uncommon event. To identify significantly higher revision rates in failing devices, large sample sizes with patients followed-up over a long-term are required[204]. Requiring large sample sizes results in a large number of patients undergoing invasive surgery using a failing device before being able to establish an unacceptable failure rate. This is problematic as such a significant number of patients need to be exposed just to

discover the fact that a device should not be used. Therefore, using revision as the only indicator of a device's success potentially lacks the sensitivity required to identify failing implants early and in turn impacts a registries ability to do the same.

These failings of registries and other bodies such as ODEP using revision as the primary outcome measure have been highlighted with recent MoM disasters. Over one million devices have been used in patients across the world before their recall[205]. The ASR implant had an excellent early ODEP rating of 3A depicting an acceptable failure rate of less than 3% after three-year follow-up data[46]. However, this was also recalled from the market after nearly 100,000 were implanted worldwide[46,70]. Despite joint registries in use at the time, with mandatory collection of data, the detrimental effects of MoM devices "were not identified sufficiently early to protect public health"[122].

Therefore, surrogate markers are sought to identify failing devices earlier than using revision rates alone. As the systematic review presented in Chapter 2 demonstrated, RSA is the only validated surrogate marker currently available. The use of RSA as part of a phased introduction for all implants prior to market introduction has been recommended[101,113]. However, RSA alone does not "universally predict long-term"[164] and there is still a need for high-quality post-market surveillance as noted in the British Orthopaedic Association (BOA) and Medicines and Healthcare products Regulatory Agency (MHRA) joint initiative "Beyond Compliance"[101,116]. This initiative mentions the need to "identify failures at the earliest point and suggest appropriate actions". The use of registries is inadequate in its current format for post-market surveillance of hip arthroplasty implants[117,206,207] and has been for a considerable time[161].

Thus, other surrogate markers need to be identified for their potential to be used in post-market surveillance.

3.2.3. Pharmaceutical versus medical devices post-market surveillance

Post-market surveillance is not a new concept and has been employed as part of regulations for pharmaceutical industries. Pharmacoepidemiology is integral to the development and surveillance of new medications entering the UK market. Pharmacoepidemiologic studies are used to assess the need for potential medicines, evaluating pre-market approval safety and post-market surveillance. However, epidemiological studies for medical devices are rarer and more difficult to undertake leading to a potential lack of validity[189,190].

Once a device or drug is approved it will be used by a more diverse and larger population than those found in clinical trials. Traditionally, the reporting of spontaneous adverse events (AE) are used to perform post-market medical device vigilance[208]. Exactly what this entails varies depending on product category, approval type and country. For THA surgery, the main recorded adverse event is revision surgery which can be monitored by registries. Here, newly introduced implants can be compared to a clinically proven standard. However, as discussed, there are limitations to relying solely on measuring revision rate through registry data for post-market surveillance. While the use of clinical trials to compare new implants to the standard are necessary, the use of post-market surveillance using real world data to assess how the devices perform outside of a controlled environment is essential. Randomised

Controlled Trials (RCT) offer one of the highest levels of evidence to demonstrate causality and in clinical research is considered a gold standard[209]. It is challenging to ensure identically matched patients between the exposure and comparator groups in clinical research. RCTs offer a method to achieve this and therefore, potentially adjust for unknown confounders in addition to any known confounding variables. Furthermore, RCTs offer the opportunity to construct a prospective study protocol with well-defined inclusion/exclusions criteria, intervention and endpoints which is often not possible with observational studies[210]. However, there are scenarios in which conducting RCTs would be unfeasible or unethical and often clinical practice differs to the environment of a controlled trial potentially lacking generalisability[209]. In addition, used alone, clinical trials are not thorough enough to establish the safety of new devices, as they investigate the short-term. Revision is a rare outcome and an acceptable performance of a device is calculated as a revision rate of less than 5% at ten years[130]. Thus, it would be very expensive to conduct randomised controlled trials to investigate the revision rate of every new implant. Furthermore, clinical trials do not have adequate follow-up, typically two to five years, that could establish whether an implant has an acceptable ten-year revision rate. It is difficult to detect various complications through trials with a relatively short follow-up period[211,212]. Such trials may not be performed using rigorous randomisation or blinding as compared to those conducted in pharmaceutical trials. Unlike the pharmaceutical industry where there is the World Health Organisation Vigibase for drugs, none such unified database exist for medical devices[213].

Clinical trials only investigate selected patients and bias could be introduced through extensive patient selection or exclusion. Inherent limitations exist with patient recruitment resulting in interpretations that may not be representative of the general population especially with strict entry criteria. Similar limitations occur in prospective cohort studies with recruitment difficulties and large costs in data collection. Rapid innovation and advancing technologies mean that it would not be feasible to wait for such trials to finish.

A learning curve for new devices exists, and successful operations can be influenced by the experience of the surgeon[124]. Furthermore, a device may change from its original design upon market entry with incremental changes occurring over time[124]. Finally, a surgeon is exposed to a significant number of new devices all of which have limited indications for clinical use compared to the introduction of a single novel drug[214]. Evaluating medical devices requires incorporating these unique factors. Therefore, observational studies using registry data offers an opportunity for triangulation of evidence in conjunction with pre-market clinical trials to ensure generalisability and long-term follow up for safety outcomes.

Current post-market surveillance is considered passive and reactive in nature calling into question its quality[213]. In a recent study, nearly 45% of adverse events reported were classified as "unknown" making post-market analysis challenging[213]. One of the contributing factors highlighted in this study was the lack of completeness of reports concerning adverse events with device samples not being sent for evaluation.

Therefore, there is an urgent need for post-market surveillance using epidemiological methods as it "is necessarily more observational than the controlled experiments"[128]. Post-market safety surveillance using routinely collected data on a large scale is essential to ensure rapid innovation while maintaining patient safety.

3.2.4. Potential alternative surrogate markers

Chapter 2 highlighted the potential for using Patient Reported Outcome Measures (PROMs) using the Oxford Hip Score (OHS).

OHS utilises a set of 12 questions asking patients to describe pain and function in their hip during the previous four weeks, from which an overall score is given and was first introduced in 1996 in an attempt to measure morbidity with hip pathology[215]. Each question is on a Likert scale with values from 0 to 4. An overall score is created by summing the responses to each of the 12 questions. The total score can range from 0 to 48, where 0 is the worst possible score (most severe symptoms) and 48 is the best (least symptoms). OHS has been used in the clinical setting for considerable time[216] and has been validated in multiple studies[217]. Furthermore, OHS has correlated with patient satisfaction and other measures of health related quality of life such as the Euroqol 5D[218]. The need for such measures followed a shift in how outcomes after THA are assessed with increased focus on quality of life and satisfaction from the patient's perspective[219]. The systematic review of alternative surrogate markers identified one study investigating OHS showing promise[174]. A score of less than 27 measured six months post-operatively predicted higher revision rates within two-years following surgery. Although amassing PROMs data can be time consuming and associated with increased costs, it is collected routinely in the United Kingdom, thus offering a

potentially simpler and far less costly alternative to other surrogate markers such as RSA[220].

Since the study completion in Chapter 2, further investigation has been conducted investigating PROMs with long-term revision rate[221]. In this study, several PROMs were assessed at one-year following surgery and were associated with revision rates at ten years. PROMs data investigated included OHS, EQ-5D questionnaire and a visual analogue scale. The strength of associated was determined to be moderate with an area under the curve of 0.68. Therefore, PROMs may have a role in the future as a surrogate marker of long-term outcome for primary THA in post-market surveillance. As recognition of the limitations for using revision as the sole indicator of outcome in THA surgery, and for the potential of using PROMs data, PROMs has been combined with registry data in the early 2000s[222,223]. However, there are limitations with using PROMs data. There are multiple factors that can influence revision risk in THA such as BMI[224]. Only age, sex and post-operative PROMs were included as potential confounders within this study, and only a moderate association was found[221]. Thus, PROMs data may not be able to predict long-term revision rate or be powerful enough to detect failing implants early.

Clarity of questions used in PROMs has been highlighted as a pitfall with a third of patients being unsure how to answer[219]. Thus, the perceived ambiguous nature of some questions can lead to bias and confound interpretations, especially as these questions can be omitted and not counted in the final score[225]. Double-barrelled questions also contribute to question ambiguity potentially leading to bias[226]. A further difficulty identified by patients is the lack of ability to quantify the intensity of pain as this can be

influenced by other factors such as activity levels or medications[219]. Again, nearly one in four patients could not give an answer for their "average" level of pain in the preceding four weeks. Analgesia can mask the intensity of pain or patients have adapted to endure it and thus struggle to answer questions relating to their "average" pain[227]. For some questions, patients may be unable to complete one answer as they did not feel their situation fits into a singular category[219]. When calculating the final score clinicians have used the highest answer, but this can lead to bias[174].

Patients found it challenging to differentiate between disability due to the implant and as a consequence from co-morbidities[228]. Patients with higher scores were associated with having multiple co-morbidities[229]. Therefore, if a patient is unable to distinguish between disability or pain arising from the hip device versus other co-morbidities, PROMs can lose their specificity and would be unsuitable as a surrogate marker of implant performance[219]. However, this could influence the indication for surgery and the decision for primary or revision operation. This may also be an indication that joint replacement may have been inappropriate in such patients without considering a holistic approach and would explain why some patients have poor outcomes. When multiple factors exist that can influence responses there can be significant floor and ceiling effects[230]. Therefore, there is great importance to adjusting for potential confounding variables which produce opposed results when comparing adjusted versus unadjusted results[132].

Registry data is rich and varied with the potential to be a powerful tool for post-market surveillance but is under-used by researchers[231]. Therefore, there is a need to investigate the potential of other surrogate measures to be used in post-market

surveillance of THA devices. Such surrogate markers should be routinely collected, in large numbers and represent symptoms measured objectively over time rather than within the last four weeks as with PROMs such as the OHS.

3.2.5. Analgesic utilisation within the first post-operative year following Total Hip Arthroplasty as a surrogate marker of device survival

THA devices were designed to treat Osteoarthritis (OA)[3], a clinical syndrome of arthralgia with reduced functional ability leading to impaired quality of life[232]. One of the leading causes of pain worldwide is OA with over half of elderly patients affected[233]. Pain is the main symptom resulting in impaired function and is the driving force behind patients seeking medical advice[234,235]. Conservative interventions for managing OA symptoms include weight loss, activity modification and the use of walking aids[236]. However, the principle treatments for OA are the management of pain through pharmacological means followed by joint replacement surgery if end-stage disease is present[237,238].

The Darzi report has illustrated the importance of measuring outcome from the patient's perspective[239]. THA revision surgery is not a complete marker of success. A device may be classified as successful if it survives for 15-years[202], but a patient may disagree should they have experienced persistent pain and disability. Persistent pain following primary THA surgery can be an indication of post-operative complication[240–243]. One of the main indications for patients undergoing primary THA is pain[244,245] and

therefore, pain can be regarded as a vital indicator of successful surgery[243]. Patient Reported Outcome Measures (PROMs) are increasingly used to measure successful outcome in THA, evaluating pain among other measures[246–249]. PROMs offers the potential to measure pain as an outcome of THA surgery[202]. Studies investigating outcomes using PROMs have demonstrated 10% to 15% of patients who have undergone primary THA surgery report persistent pain[250]. Other studies estimate post-operative pain can occur between 7% to 23% of patients[251]. However, analysis of revision rate alone fails to identify patients with persistent pain or disability. Therefore, there have been calls for methods to measure pain that can subsequently be used in conjunction with revision rates for accurately monitoring outcomes in THA surgery[202]. In addition to being an indication for primary THA, pain is one of the commonest reasons for revision hip arthroplasty[11]. Several studies have shown revision surgery occurring due to persistent pain alone[174,243,252].

One study demonstrated the association between PROMs measured within the first post-operative year of primary THA and subsequent revision rate[221]. Despite its potential PROMs may not be a suitable surrogate marker due to several limitations. The reliability and validity of PROMs can vary depending on which is used in which context; there can be significant implementation costs especially for their use as part of a national database; there can be challenges with integrating its use within current workflow systems; and there are potential difficulties obtaining adequate patient participation which could lead to analysis issues when used as an outcome[246,253].

However, pain could be a vital measure of outcome. There was a higher incidence of pain for patients who had Metal-on-Metal (MoM) or resurfacing hip arthroplasty[11,72]. Pain alone accounted for more than one in five patients as the indication for revision surgery in MoM implants[254]. These device bearings have had significant failings and highlighted the need for alternative early surrogate markers of outcome[46]. Therefore, monitoring pain may be a vital early indicator of failing implants in large cohorts and a useful endpoint of outcome[174,255].

Aghayev et al showed pain intensity follows an almost parallel course with aseptic loosening for devices subsequently revised in THA surgery[256]. In the study by Eneqvist et al, pain in isolation was shown to have the highest predictive power for subsequent revision in primary THA with an area under the receiver operator curve of 0.65[221]. Pain is generally considered the principal indication for THA surgery, it has been shown to predict revision and is a better discriminator between devices compared to revision as an endpoint in survival analysis[255]. Furthermore, the incidence of pain is far higher than revision, allowing for any methods used to detect differences in devices to be detected both in a smaller cohort of patients and earlier than using revision alone[255]. This concept is vital when assessing new devices during post-market surveillance as they have substantially lower revision rates initially[255]. PROMs, such as the Oxford Hip Score, measure pain over the proceeding four weeks prior to the completion of the survey. However, pain is not static but dynamic[219] and symptoms can significantly fluctuate over time with multiple influences[257]. Therefore, other surrogate measures of pain over a prolonged period are needed to potentially be used as early signals of device survival.

Measuring analgesic use following THA surgery can potentially be an alternative marker of persistent post-operative pain[258]. Analgesic use increases prior to surgery peaking in the immediate peri-operative period and decreases within the first year as THA is a treatment for pain[259]. Therefore, it is conceivable that persistent pain could indicate problems with the device. Pharmaceutical records can be used as a measure of analgesic use and therefore as an indicator of persistent post-operative pain[258], as they meet several conditions[243]. Firstly, these medications are commonly used to treat musculoskeletal pain[260]. Furthermore, analgesic medication recorded in pharmaceutical data are only obtained upon recommendation by the doctor and thus used when significant pain is present[243]. Pharmaceutical data is routinely collected and could provide an efficient and cost-effective method for assessing pain to be used as early signal surveillance in large patient cohorts to monitor device success[243].

One of the analgesic groups regarded as the "mainstay of the pharmacologic management" and are "strongly recommended" for patients with musculoskeletal pain is Non-Steroidal Anti-Inflammatory Drugs (NSAID)[260]. Nearly 60% of patients use NSAIDs to manage musculoskeletal pain[261], and patients suffering from OA prefer using NSAIDs due to improved pain relief[262]. In general, NSAID are some of the most extensively used medications across the world with strong evidence in their treatment of pain and inflammation[263]. Their mode of action is by inhibiting cyclooxygenase, which is a pro-inflammatory enzyme[263]. NSAID are recommended in the management of musculoskeletal pain when simple conservative methods or simple analgesics such

as paracetamol are not sufficient[264]. They are also regarded the initial choice of oral treatment for osteoarthritis[260].

NSAID use has been implicated as a marker of pain in hip devices[265]. While medication such as ibuprofen is readily available over the counter in most countries, higher doses, and other preparations of NSAID (such as Diclofenac or Naproxen) are only available through prescription. Prescribed NSAID indicate increased severity of pain and thus likely to indicate significant problems with a device[243]. Therefore, NSAID can be used as a surrogate measure of pain and thus a potential surrogate marker of THA outcome.

In addition to NSAID, one of the most commonly used analgesic medications for musculoskeletal pain are opioids[266]. Over one third of patients use opioids to manage musculoskeletal pain[261]. Their mode of action is by binding to μ opioid receptors which are expressed along the pain pathway[267]. Opioid medication is used for the treatment of moderate to severe pain in musculoskeletal conditions[243], and is commonly prescribed in the post-operative period following THA[268]. They are usually called upon when other analgesic medications have failed or are contra-indicated due to reasons such as co-morbidities or intolerance, or the patient has refractory pain[260,269,270]. The American College of Rheumatology guidelines recommend Tramadol use along with oral NSAIDs or intra-articular corticosteroid injections as first-line treatment of hip OA pain[271]. There are concerns with adverse events and prolonged post-operative use[236]. Higher doses of opioid medication is associated with increased use one year post-operatively[272]. Furthermore, there have been questions regarding the use of opioids due to their minimal benefits compared to placebos[273]. Despite these factors they are

used widely with nearly 25% of patients taking opioids in the first-year post THA[243,258]. Interestingly, the predominant prescribers of opioid medication are general practitioners covering nearly half of prescriptions at the one year mark following THA surgery[236,274]. While the reason for this is uncertain, primary care doctors are more available to patients, especially beyond the initial post-operative period, and opioids may be prescribed to satisfy patient expectations[236]. Multiple types of opioids are available. Tramadol has been specifically recommended in the management of musculoskeletal pain in osteoarthritis of the hip as well as other joints[260]. Tramadol is on the World Health Organisation analgesic ladder and is considered a 'weak opioid'[275]. The mode of action is a combination of μ opioid receptor agonist and inhibition of serotonin and norepinephrine reuptake[276]. The benefits of Tramadol, like other opioids, are found to be modest in the longer term up to one year[277]. It has been demonstrated to reduce intensity of pain and improve function in patients suffering from OA to a similar level as diclofenac[278]. Although there are some concerns of a potential risk of dependence there are no differences in prevalence of dependence compared to NSAIDs[279]. For other 'non-tramadol' opioids, authorities actively recommended against their use in patients with osteoarthritis of the hip or knee particularly for long-term use[260]. Furthermore, other opioid medications, such as buprenorphine and fentanyl, are more likely to be used for pain caused by other reasons and have been excluded in other studies[266]. Therefore, tramadol is recommended over other non-tramadol opioids[260].

Alternative types of analgesic medications could potentially be used. Moderate to severe pain is associated with revision[221,280] and would subsequently be more indicative of issues with an implant. Despite advancement in modern medicines, treating moderate to severe pain remains challenging[271]. A number of reasons could contribute

including the natural progression of OA, the heterogeneity of patients, interactions between co-morbidities and medications, the efficacy or safety profiles of analgesics available, and multiple mechanisms for pain[281]. The underlying cause of pain in OA is poorly understood with complex theoretical mechanisms[232]. Currently there is no objective method to measure pain in musculoskeletal disorders. Measuring bone or cartilage damage using imaging modalities such as radiographs, ultrasound or magnetic resonance imaging has proved ineffective at predicting either the presence or severity of pain in patients with OA[282]. Approximately 40% of patients with damage identified on imaging experience no pain yet patients with no or minimal damage experience debilitating pain[283]. Traditional understanding of osteoarthritic pain considers the main mechanism is the nociceptive pathway as result of stimulation of joint nociceptors and activation of nociceptive processes within the brain stem, spinal column and thalamocortical system[232]. However, cartilage within a joint is avascular consisting of aneural tissue and therefore, processes such as joint inflammation, bone damage and central pain perception could play vital roles[284]. Other findings such as specific expressions of neurons suggest a neuropathic component to pain in OA patients[232]. Several studies have confirmed that pain in OA is not exclusively via the nociceptive pathways with approximately 20% of hip pain presenting with neuropathic like symptoms[285–289]. One study revealed the prevalence of such symptoms for patients with hip OA is as high as 37%[289]. Other studies have found almost a third of patients with osteoarthritis have a neuropathic element and such patients would benefit from neuropathic medication[232,290]. These symptoms could include paraesthesia, numbness or hyperalgesia among others[291]. However, conventional analgesia used first-line in the management of OA target nociceptive pathways and therefore are limited in their effect of treating neuropathic like symptoms[289], and thus ineffective at treating

persistent pain. It is important to note that radiographic severity is not associated with neuropathic like symptoms, suggesting that such symptoms occur due to a "peripheral or centrally augmented pain state" as a result of "generalised central sensitisation"[289], rather than cartilage degeneration. One theory suggests this central sensitisation stems from long-term constant joint nociceptive sensitisation[292] and thus neuropathic pain in these patients is likely due to OA rather than other reasons. Furthermore, older patients are more likely to have co-morbidities that preclude them from having first-line analgesics for OA due to side effects or drug interactions[271]. Therefore, there are management concepts calling for the need to create bespoke conservative OA management in patients experiencing neuropathic like pain using neuropathic analgesic medication[259,293]. Medication used to treat neuropathic pain are used in the management of musculoskeletal pain and a significant proportion of patients with persistent post-operative THA pain may be treated with these neuropathic analgesics[251,266]. Persistent pain in the year following THA surgery may result in increased rates of neuropathic medication use[266,294]. The presence of neuropathic pain prior to hip replacement seems to be rare with such patients instead describing their pain in conjunction with an inflammatory factor[295]. Therefore, the use of neuropathic analgesic drugs could be an indication of severity of pain requiring stronger control. One study found approximately 60% of pre-operative users of neuropathic pain medication continued to require them one-year post THA surgery[266]. Furthermore, 40% of neuropathic analgesic users at one-year post-operatively did not have these medications prior to surgery[266]. This suggests there is the potential for a neuropathic component to pain following hip arthroplasty that is not dependent on prior use.

Atypical analgesic medications are those used in the management of pain but their primary use is for conditions other than pain[296]. Commonly used atypical analgesics are centrally acting medication for the treatment of chronic or neuropathic pain[260] including medicines that are used for depression or epilepsy as their primary treatment condition[297]. Guidelines from the National Institute for Health and Care Excellence (NICE) recommend the first-line medication used in treating neuropathic pain are Duloxetine, Amitriptyline, Gabapentin and Pregablin[298]. Neuropathic pain medications are also used following THA surgery with a prevalence of 3% at two years post-operation[266]. Furthermore, an increase in use was found when compared to pre-operative neuropathic pain medication use suggesting an indication is persistent pain[266].

There are two main types of antidepressants used in treating neuropathic like pain for musculoskeletal pain: Tricyclic Antidepressants (TCA) and Serotonin-Norepinephrine Reuptake Inhibitors (SNRI)[296]. Antidepressants are commonly used with up to 3% of all prescriptions used in the treatment of musculoskeletal pain[299]. However, despite the majority of studies investigating analgesic effects in arthritis using neuropathic analgesia showing over a 50% improvement in approximately 25% to over 60% of patients, these studies have significant limitations and small sample sizes[300]. SNRIs inhibit reuptake of serotonin and norepinephrine selectively[271]. These medications provide less analgesic effects but are better tolerated compared to TCAs[271], and are thus not recommended first-line for the treatment of arthritic patients[301]. Duloxetine is the only centrally acting medication that has an adequate evidence base for its use in musculoskeletal pain to allow for formal recommendations[260], and has been shown to

be effective for patients with OA[302]. TCAs are commonly used to treat neuropathic pain and are in significant distribution[303]. Their mode of action is to inhibit the reuptake of serotonin and norepinephrine[304]. TCAs have been demonstrated to significantly improve pain in arthritic patients[301,305,306] aside from anti-depressive effects. Part of this benefit is potentially from improvements with sleep and fatigue in such patients[307]. A first-line treatment for neuropathic pain, and thus subsequently neuropathic-like pain, is Amitriptyline which has many decades of successful treatment in patients[297].

The second type of medication widely used to treat neuropathic pain in conjunction with antidepressants are anticonvulsants[271]. Examples of anticonvulsants include Pregabalin and Gabapentin which are commonly used medications for chronic pain[260], and are recommended as first-line in the non-specialist setting[298]. The mode of action includes modulating the release of neurotransmitters such as serotonin and noradrenaline[271], but the analgesic mechanisms remain unclear[308]. In doing so, anticonvulsants can provide an analgesic effect for patients suffering with pain due to central sensitisation[271]. Despite no definitive evidence for their use in arthritic patients, these analgesics are commonly used in musculoskeletal pain[260]. Pregabalin functions on voltage-dependent calcium channels and subsequently reduces neurotransmitter release[309,310]. Pregabalin has been shown to be effective in reducing pain in a preclinical study[311]. Furthermore, Pregabalin can reduce the consumption of opioid medication following THA surgery[312,313] and may be more favoured in patients for which opioid medication is contraindicated, who have intolerable side effects or to reduce the need for long-term use. Gabapentin is a third-generation medication[314] that has effects on both peripheral as well as central pain activity[315]. It may be used post-operatively in

THA surgery in patients where combined use of NSAIDs and opioids would have undesirable gastrointestinal or renal function side effects[316]. Gabapentin has shown a borderline significant analgesic effect in the treatment of persistent post-operative pain[317].

3.2.6. Healthcare utilisation within the first post-operative year following Total Hip Arthroplasty as a surrogate marker of device survival

One alternative data source for surveillance that has been proposed is the use of Electronic Medical Record (EMR) systems[318,319]. A recent systematic review has demonstrated at least three in four hospitals in the United Kingdom use EMR systems[320] and there is nearly universal use in primary care since 1996[321,322]. The widespread adaptation of EMR systems leads to new opportunities with readily available national data on patient encounters which has already been shown to improve understanding and treatment of diseases[323,324]. Using EMR systems for surveillance is not a new concept and has been proposed[318]. A wealth of patient encounter data can be available which is not limited to the specific data collected in joint registries such as inpatient or surgical encounters. When using primary care databases, available data can expand to all types of encounters and potentially any contact a patient may have within the healthcare system[325]. The scope of information is now widely available with relatively easy accessibility and is now becoming a reality for researchers to investigate[325]. In addition, data is available in real-time, potentially ensuring earlier identification of device problems[319]. EMR systems provide the

opportunity for "timely surveillance of failures" and the early identification of failing Total Hip Arthroplasty (THA) devices[319].

Currently, monitoring of THA devices is based on revision rate as an endpoint. As revision can be influenced by several patient characteristics, revision is not straightforward to interpret. Therefore, expansion of outcome measures is required and one such indicator is healthcare resources utilisation[231]. Data on encounters or healthcare services utilisation could potentially be used as early "indicators of poor surgical outcomes"[325] and therefore be used to identify failing THA devices. The potential use of routinely collected data, such as analgesic use, as a surrogate marker of pain and THA revision rate has been demonstrated in previous chapters and in other studies[243,258,326]. Post-operative care has been demonstrated to have an association with subsequent revision rate following primary THA surgery[319]. Therefore, data such as the frequency of General Practitioner (GP) visits and referrals to either the Orthopaedic or Physiotherapy departments can be used as a possible measure of outcome.

The frequency of GP consultations by patients with musculoskeletal problems is substantial[327]. A significant proportion of patients with lower limb musculoskeletal complaints have contact with their GP accounting for nearly a third of all consultations[328], and almost half of those suffering will seek consultation with their GP[329]. There are indications that patients with increased severity of Osteoarthritis, in both pain and function, have an increased risk of visiting their doctor. Patients with more pain, according to WOMAC scoring, are more likely to seek medical help from their GP compared to patients who did not seek advice[330].

The number of GP visits has been demonstrated to be significantly increased in those patients requiring THA surgery compared to those patients who do not suffer with Osteoarthritis[331]. Patients who subsequently undergo hip replacement require an additional 1.8 more visits to the primary care physician within a year prior to surgery[331]. Furthermore, nearly all the consultation costs for patients with Osteoarthritis who undergo THA surgery was due to GP consultations[331]. This same study concluded that monitoring resource use could be a "predictor of need for" THA surgery[331].

Measuring the frequency of encounters as a potential indicator of surgical outcome has been demonstrated in other areas of Orthopaedic surgery[325]. In this study, patients with more frequent visits to either the Orthopaedic outpatient department or primary care physician was associated with a significant increase in risk of revision surgery[325]. Furthermore, one study investigated the use of healthcare utilisation to predict subsequent revision rate and found patients requiring at least seven visits to their primary care physician had over a two-fold risk of early revision compared to those needing zero to two visits[319]. In addition, another study demonstrated patients with six or more visits to a primary care physician in the same post-operative period are over four times more likely to require subsequent revision surgery[332]. Therefore, all-cause frequency of visits to the primary care physician may be a potential early indicator of failing devices.

In the UK patients do have not have direct access to secondary care. GPs are the gateway to secondary care with referral required from a GP before a patient is able to see a specialist[333]. Musculoskeletal conditions are among the most common diseases

patients are referred to a specialist from primary care[334]. In principle, any referral to secondary care can only be made by a GP whether in the national or private healthcare systems. GPs can determine a patient's eligibility for consideration of THA surgery[335], and potentially be a good representation of the proportion of patients seeking medical care[327]. However, there is no definitive criteria of how a decision for referral is made[336,337], with pain and limited function being the predominant clinical predictors of patients being added to the waiting list for THA[338]. Despite this, the implication for referral to secondary care is that there is a significant problem. One study demonstrated no difference in pain severity for patients seen in the community setting compared to those referred to Orthopaedics[330]. However, patients with more clinical signs were more likely to be referred to a specialist department[330].

The National Institute for Health and Care Excellence (NICE) recommend referral for consideration of joint replacement before there is severe pain or prolonged functional impairment[339]. Therefore, primary care physicians may refer a patient with significant and prolonged pain following THA surgery potentially indicating an underlying problem with the device, and thus be an early indicator of poor outcome. Half of patients referred to the Orthopaedic department by the GP for hip Osteoarthritis go on to have a joint replacement within one year[335]. Furthermore, almost 20% of patients are recommended for replacement but scheduled later[335]. Therefore, approximately 70% of patients being referred by their GP to the Orthopaedic department, subsequently require THA surgery. In addition, worsening pain increases the likelihood of being referred and subsequently placed on the waiting list for surgery[335]. There remains a small proportion of patients, approximately 10%, who will refuse surgery if offered due to personal preference[335]. It is important to note that not all patients referred will undergo joint replacement. Reasons for refusal are multifactorial and can include

previous negative experience and healthcare professional's opinion reinforcing a patient's decision[340].

The frequency of visits to the Orthopaedic department within the first post-operative year following THA surgery has been investigated. Patients who have attended the Orthopaedic department six or more times within four months following THA surgery are over 15 times more likely to need revision surgery[332]. In another study investigating healthcare utilisation and revision rate in THA surgery, patients requiring at least six visits to the Orthopaedic department had more than a four-fold increased risk of revision compared to those needing zero to three visits[319]. Due to the longevity and prolonged use of THA devices in healthcare, the routine post-operative follow-up in the Orthopaedic department is markedly less than those found to be associated with later revision. The British Orthopaedic Association (BOA) recommend an initial early post-operative review following THA surgery within eight weeks[341]. They also recommend long-term post-operative follow-up at one year, seven years and three yearly thereafter for those devices with an Orthopaedic Data Evaluation Panel (ODEP) 10A rating[341]. There have been calls for 10A rated devices to be discharged following the first post-operative review in the presence of an adequate referral system[13]. For unproven devices or those newly introduced to the market, the BOA recommends annual follow-up in the first five years, two-yearly until year ten, and then three yearly thereafter[341]. NICE guidelines recommend referral to the Orthopaedic department for any patient developing new or worsening pain, or loss of function due to their joint replacement[342]. Therefore, those patients referred within the first year require more visits than recommend and may be an early indicator of failing devices.

As with all secondary care, referral from a GP is required for patients to be seen by a Physiotherapist[333]. Physiotherapy referrals are dominated by patients suffering from musculoskeletal issues[343]. NICE guidelines recommend patients should be offered non-surgical options before any consideration of joint replacement including Physiotherapy[339]. Approximately one third of all patients suffering with Osteoarthritis are referred to the Physiotherapy department[335], representing a substantial method for managing pain in these patients. Therefore, it is possible GPs will refer patients to the Physiotherapy department before considering referral to the Orthopaedic department should ongoing problems, such as pain, occur post-operatively. For Osteoarthritis, GPs refer patients to Physiotherapy more readily than to the Orthopaedic department and potentially reduce the number of referrals to Orthopaedics[344].

Patients require rehabilitation following THA surgery with step-wise increase of activity over several months[345]. The BOA recommend significant Physiotherapy team input during the immediate post-operative period until discharge[341]. Thereafter, there are no recommendations for follow-up with the Physiotherapy department but must remain contactable should any problems occur[341]. NICE guidelines also give recommendations for the Physiotherapy team to give patients advice on self-directed rehabilitation prior to the patient leaving hospital[342]. Patients who are finding difficulties with managing activities of daily living or have ongoing functional impairment may need supervised input[342].

One study investigated the frequency of Physiotherapy outpatient visits post-operatively and demonstrated an association with revision rate in THA surgery[319]. Patients with three or more visits had almost a two-fold increased risk of revision compared to those requiring zero to two visits[319]. This risk increased further with

patients who required at least nine visits to the Physiotherapy department[319]. Therefore, the need for Physiotherapy input following THA surgery during the post-operative period may be an early indicator of problematic devices.

The main aim of this experimental chapter is to determine whether analgesic or healthcare utilisation could be used as an early indicator for THA success when monitoring large, combined registry and primary care data. To achieve this, these experiments aim to investigate the association between analgesic use within the first post-operative year following primary THA and subsequent revision rate.

3.3. Methods

3.3.1. Aim

The aim is to investigate the association between novel potential surrogate markers measured within the first post-operative year and subsequent revision rate in primary Total Hip Arthroplasty surgery.

3.3.2. Study design

Cohort studies.

3.3.3. Data source

Data from the Catalan Arthroplasty Registry (RACat) that is approximately 85% linked to Spanish primary care records and pharmacy invoice data from the Sistema d'Informació per al Desenvolupament de l'Investigació en Atenció Primària (SIDIAP) Database (www.sidiap.org)[346].

The RACat registry was created in 2005 following a collaboration agreement of the Catalan Society of Orthopaedic Surgery and Traumatology (SCCOT), the Catalan Health Service (CatSalut) and the Agency for Health Quality and Assessment of Catalonia (AQuAS)[347,348]. It is the first joint registry in Spain designed to assess device performance opening the way for benchmarking and post-market surveillance[349]. AQuAS is the body coordinating data collection, management and analysis[350]. RACat was formed to contribute to the improvement of health-care quality by offering health professionals and providers feedback on the outcomes of arthroplasty surgery. RACat aims to gather socio-demographic, clinical, implant and fixation-related data on all patients undergoing any type of knee or hip arthroplasty in any of the public hospitals network in Catalonia (Northeast Spain), as it has been previously described. Currently, 53 from 61 public hospitals performing hip and knee arthroplasties are sending information to RACat which represents 80-85% of the activity taking place in Catalonia. RACat has improved coverage of THA's performed in Catalonia from over 24% in 2005 to over 70% in 2010[348]. Most recent publications show coverage has increased from 80% to 85%[231]. Data quality (coverage, informed laterality and identified prosthesis) has also improved since 2005[348,349]. Completeness of data for RACat has been calculated against mandatory information collected at a national level (Minimum Basic Data set at hospital discharge) and has been determined to be 90% for

primary THA and 70% for revision surgery[351]. RACat is managed and funded by the Department of Health[349].

SIDIAP contains primary care medical records for more than five million people (approximately 74% of the population) in Catalonia, Spain[352]. Data is collected in real-time by over 3,400 General Practitioners (GPs) and primary care nurses[352]. It comprises of clinical and referral events as registered in electronic medical records, comprehensive demographic information, prescription and, corresponding pharmacy invoicing data, specialist referrals, primary care laboratory test results, hospital admissions, and, their major outcomes[353]. This information is gathered using ICD-10 codes, and structured spreadsheets designed for the collection of variables relevant for primary care clinical management, such as country of origin, gender, age, height, weight, body mass index, smoking and drinking status, blood pressure measurements, and blood and urine test results. Only GPs who achieve quality control standards can contribute to the database[354]. Encoding personal and clinic identifiers ensures the confidentiality of the information.

The SIDIAP data has previously been used in several epidemiological studies[355–360] and a number of outcomes have been successfully validated in data extracts from this source including the most common indications for Total Joint Arthroplasty: osteoarthritis[361], fractures[362], and inflammatory arthritis[363,364].

3.3.4. Ethics

Data was received following submission and approval of a protocol to the Registry committee. Idiap Jordi Gol from the CEIC (Comite Etic d'Investigacions Cliniques in Catalan language) has approved this study.

No written informed consent was obtained. Patient records and information were anonymised and de-identified prior to analysis.

The RACat uses data collected electronically by hospitals that are linked with a unique insurance identification number given to all individuals living in Catalonia[349]. The registry is stored on a desktop computer, password protected and in a locked room within a building requiring card approved entry.

3.3.5. Study population

All participants registered in both SIDIAP and RACat undergoing primary Total Hip Arthroplasty (THA) between 2005 and July 2012 were eligible.

Patients receiving revision surgery within the first post-operative year were excluded as they are more likely to be revised due to reasons unrelated to the implant such as infection and surgical technique[365]. Furthermore, any patient who underwent revision surgery within the first post-operative year would interfere with survival data for exposures that are measured over the one-year period post-operatively potentially leading to immortal time bias[366].

Patients under the age of 40 years old were also excluded. Younger adults indication for surgery is usually severe disability and multiple joint pathology for conditions such

as acetabular dysplasia[367–369]. In addition, these patients are likely to be more active leading to increased risk of revision[370]. Thus, the gold standard measure of performance of an implant with revision rate at ten years is markedly different for young patients compared to older, "for whom ten years could represent the majority of their life-expectancy"[371]. Survivorship in the patients under aged 40 years has been shown to be significantly poor compared to older patients[372]. Studies have shown patients undergoing THA surgery will have less survival at ten years compared to standard survival studies[373].

3.3.6. Outcomes

The primary outcome was all-cause revision hip arthroplasty from the first post-operative year up to 7.5 years following primary THA surgery. Revision was defined as any exchange, removal, or addition of a prosthesis component of the primary prosthesis.

3.3.7. Exposures

As the RACat registry is a real-world database only available exposures were investigated following a literature review of potential surrogate markers. Exposures investigated can be categorised into two categories:

- Analgesic use within the first year post-operatively:
    - Non-Steroidal Anti-Inflammatory Drugs (NSAID)
    - Tramadol
    - Atypical Analgesics (Pregabalin, Gabapentin, Amitriptyline)
- Healthcare resource use within the first year post-operatively:
    - General Practice visits
    - Referrals to the Orthopaedic department
    - Referrals to the Physiotherapy department

NSAID use

Data on NSAID use in the first year post-THA was obtained from the pharmacy invoice databases for all eligible participants. Usage was quantified using the World Health Organisation (WHO) Anatomical Therapeutic Chemical Classification System with Daily Defined Doses (ATC/DDD index)[374]. The DDD index is a standardised measure of NSAID to incorporate use of different types of NSAID. The DDD index represents the average dose of each individual NSAID as used for its most common indication. For example, 100mg of Diclofenac or 1200mg of Ibuprofen is equivalent to 1 DDD[375].

NSAID use, measured as DDD, within the first post-operative year is a continuous variable. The distribution was analysed using a histogram with kernal-density (Figure 3-2).

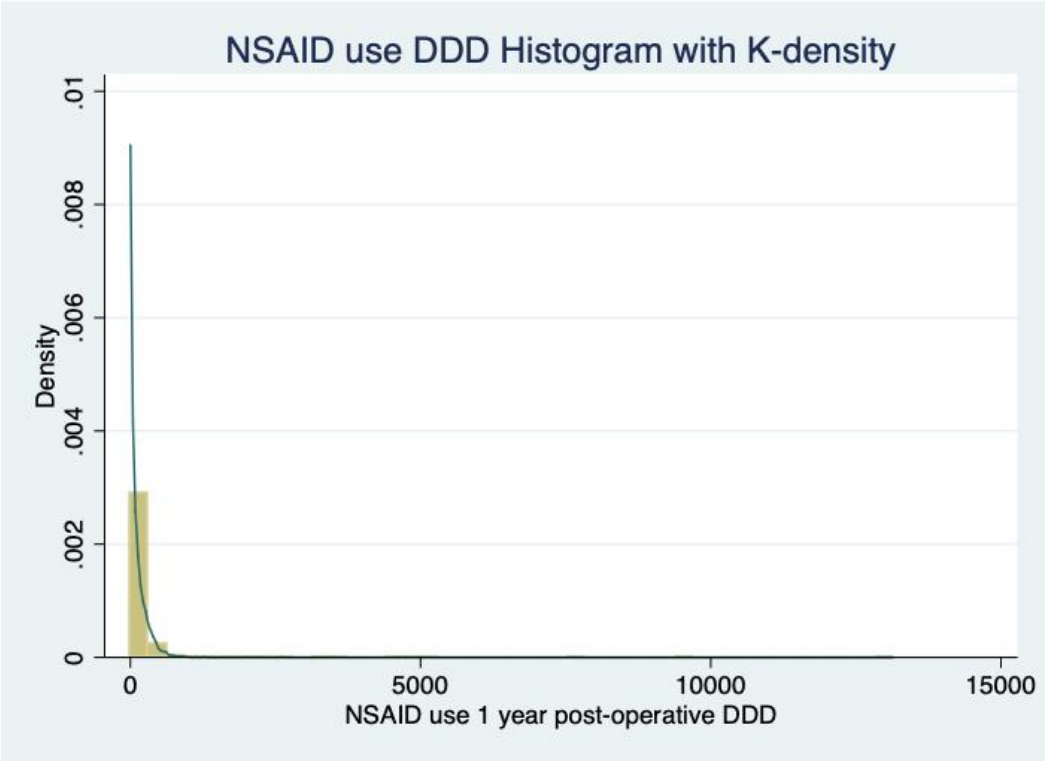


Figure 3-2: Histogram of NSAID exposure measured as Daily Defined Dose (DDD).

As the distribution is not normally distributed, assessment of histograms with transformation was made using the "gladder" code from Stata (Figure 3-3).

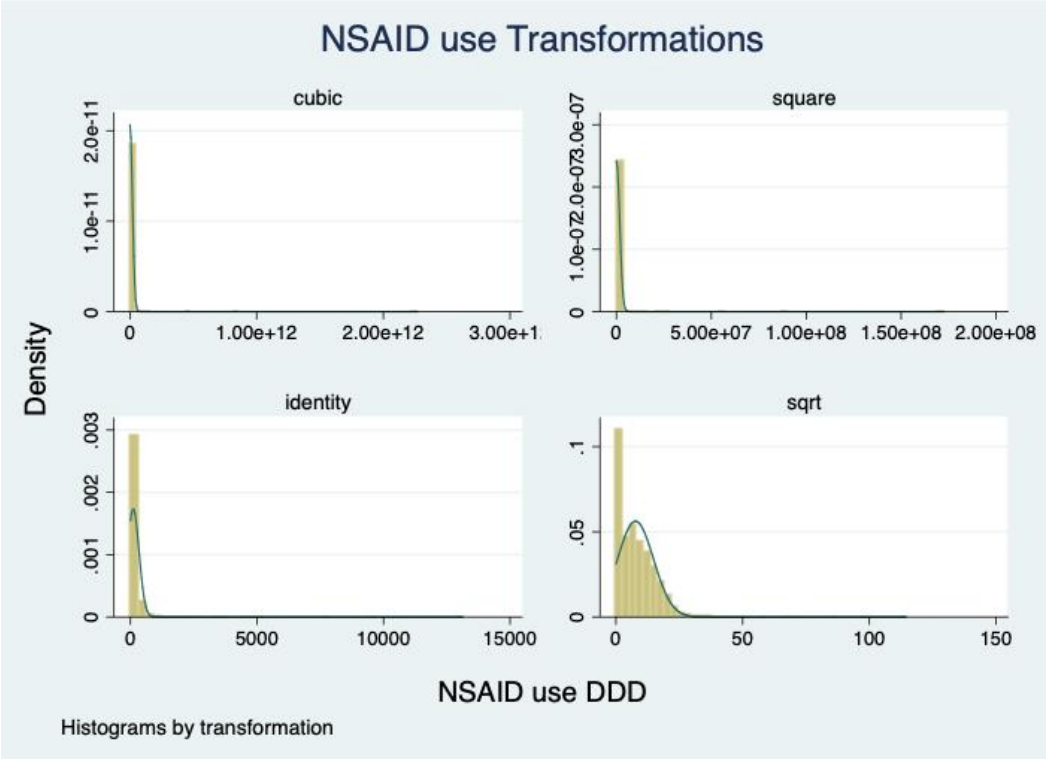


Figure 3-3: Transformations of NSAID exposure measured as Daily Defined Dose (DDD).

There were no transformations that yielded a normal distribution and thus the exposure variable was categorised into the maximum number of groups while ensuring there was sufficient power. Therefore, NSAID use was categorised into quintiles.

Tramadol use

Tramadol variables available within the database included start date, end date, and total DDD used by patients during both the pre-operative and post-operative period. However, there were no available data to calculate total DDD within the first post-operative year alone. The total DDD available covers from the first to the last

prescription overall with no method of identifying use during the post-operative period. Furthermore, the distribution of total DDD is not normally distributed (Figure 3-4).

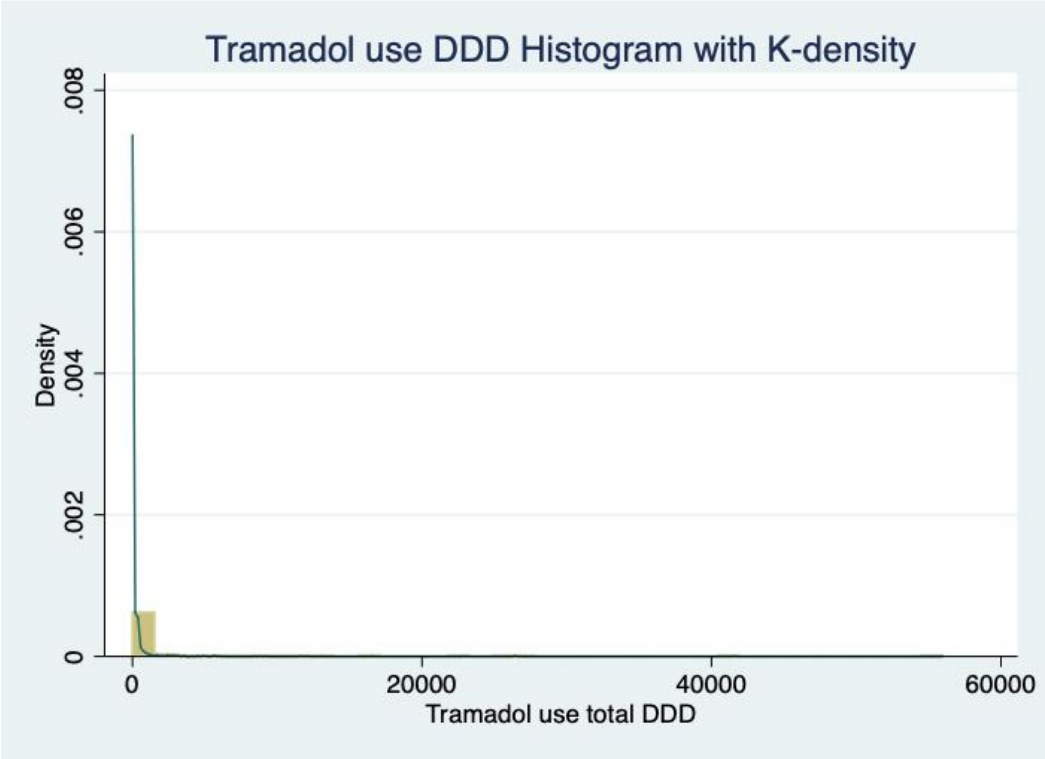


Figure 3-4: Histogram of Tramadol exposure measured as Daily Defined Dose (DDD).

As the distribution is not normally distributed, assessment of histograms with transformation was made using the "gladder" code from Stata (Figure 3-4).

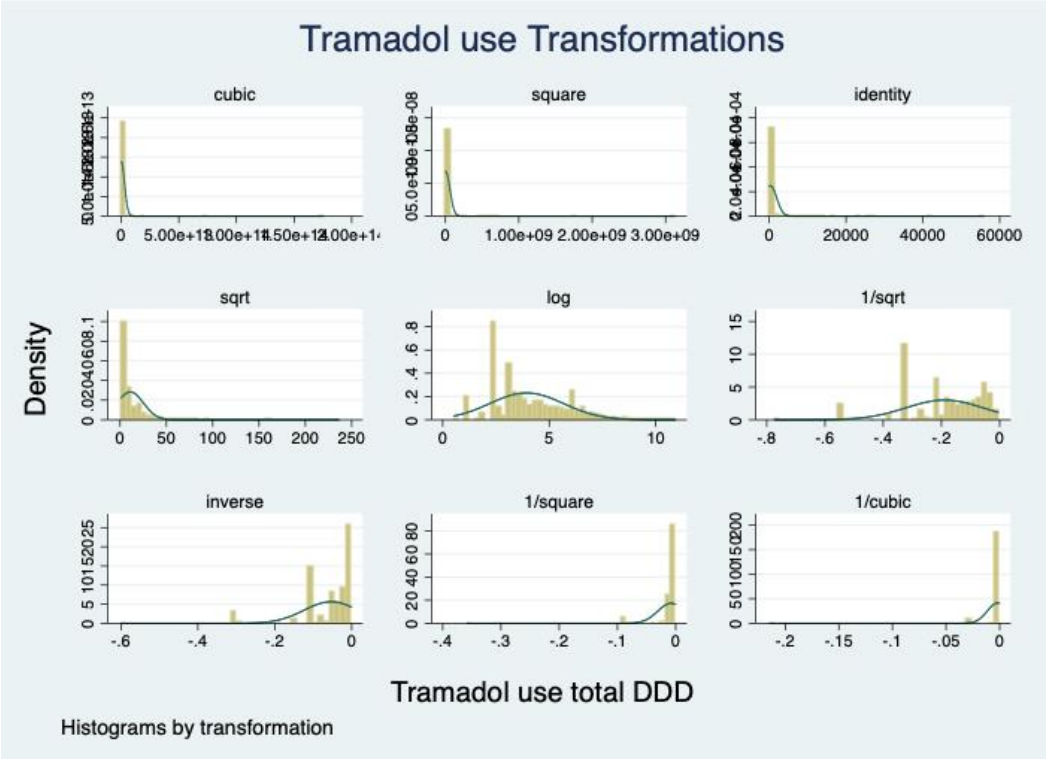


Figure 3-5: Transformations of Tramadol exposure measured as Daily Defined Dose (DDD).

There were no transformations that yielded an adequate normal distribution. Furthermore, the prevalence of patients taking Tramadol in the first-operative year within the cohort was 3,224. Any further categorisation would potentially produce groups with insufficient power. Thus, the Tramadol exposure used was categorised into a binary group of users and non-users within the first post-operative year.

Atypicals

NICE guidelines recommend Duloxetine, Amitriptyline, Gabapentin and Pregabalin as first-line medication used in treating neuropathic pain[298]. In the RACat cohort, data

was available for Amitriptyline, Pregabalin and Gabapentin only. These analgesics are the most commonly used for neuropathic pain management[376].

Indication for atypical analgesic use was not available, thus it was not possible to ascertain whether pre-users were prescribed these medications for neuropathic pain in general[297] rather than OA neuropathic-like pain[289]. Currently, there is no gold standard method to measure or detect neuropathic pain in patients with OA[289]. Neuropathic pain is different to that experienced from healthy nerves derived from damaged tissue as in the case of arthritic arthritis[297]. One study suggests patients who use medications for neuropathic pain is largely limited to those using them prior to surgery[266]. Therefore, pre-operative users of atypical analgesics were excluded from analysis as this could represent neuropathic pain rather than neuropathic-like pain as a result from OA which would represent persistent post-operative pain that could indicate a problem with the device.

Variables available within the database for three identified atypical analgesics are start date, end date, and total DDD used by patients during both the pre-operative and post-operative period. However, there was no available data to calculate total DDD within the first post-operative year for any of the atypical analgesics alone. The total DDD available covers from the first to the last prescription overall with no method of identified use during the post-operative period.

Furthermore, the distribution of total DDD is not normally distributed (Figure 3-6).

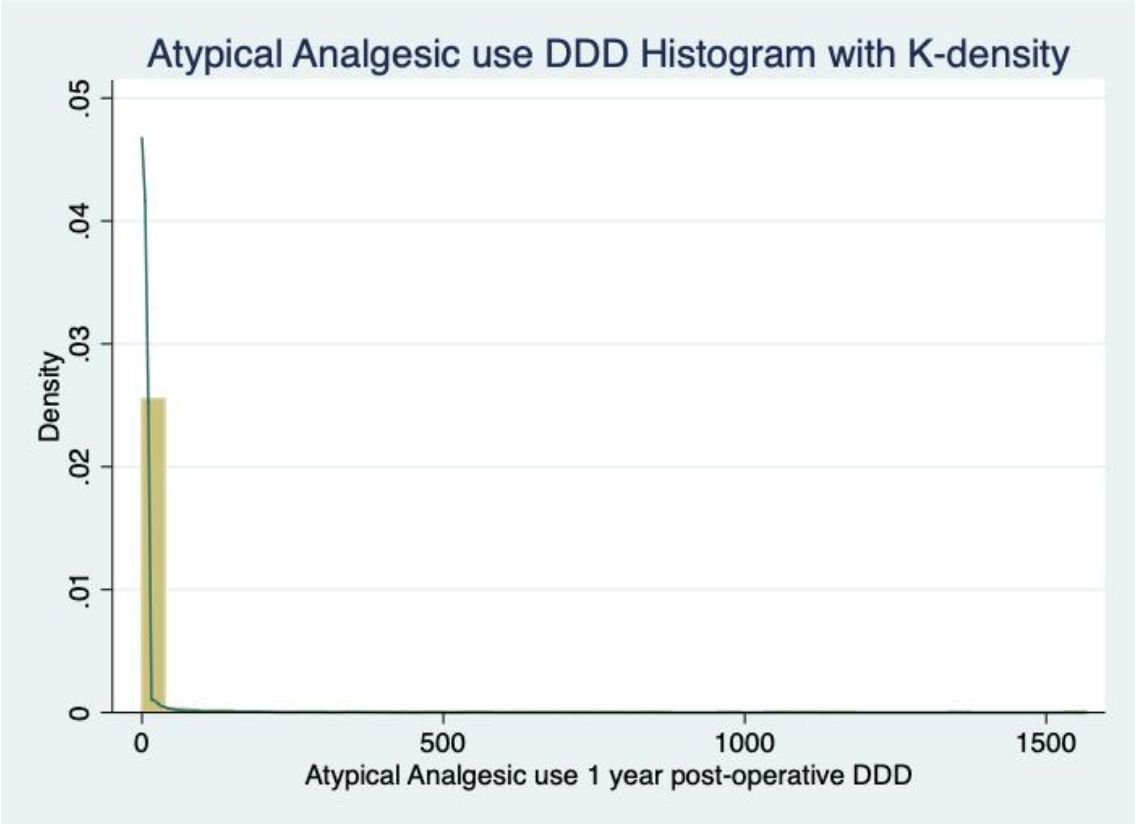


Figure 3-6: Histogram of Atypical Analgesic exposure measured as Daily Defined Dose (DDD).

As the distribution is not normally distributed, assessment of histograms with transformation was made using the "gladder" code from Stata (Figure 3-7).

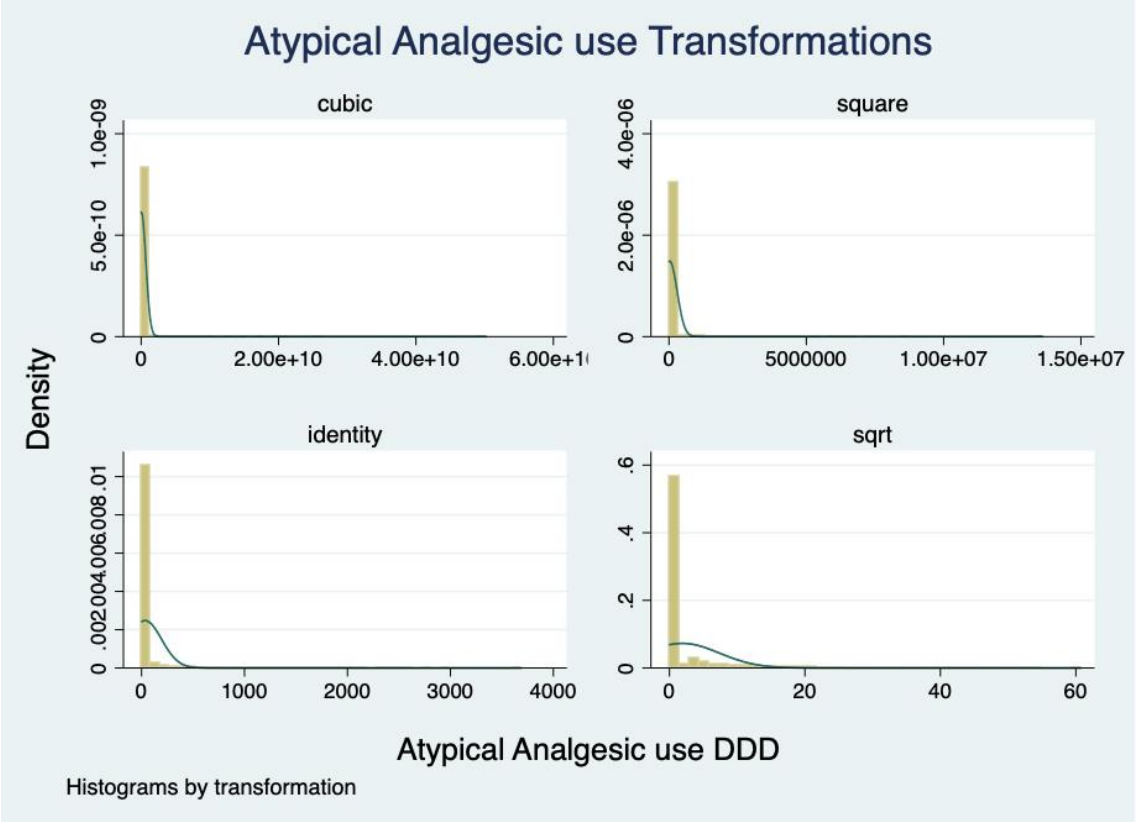


Figure 3-7: Transformations of Atypical Analgesic exposure measured as Daily Defined Dose (DDD).

There were no transformations that yielded a normal distribution. Furthermore, the incidence of patients taking Atypical Analgesics in the first-operative year within the cohort was 872. Any further categorisation would potentially produce groups with insufficient power. Thus, the Atypical Analgesics exposure used was binary between users and non-users within the first post-operative year.

Frequency of GP visits

Data was available on the number of all-cause GP visits in the first year post-THA which was extracted from the SIDIAP database for all eligible patients. GP visits, measured as a frequency count within the first post-operative year, is a continuous variable. The distribution was analysed using a histogram with k-density (Figure 3-8).

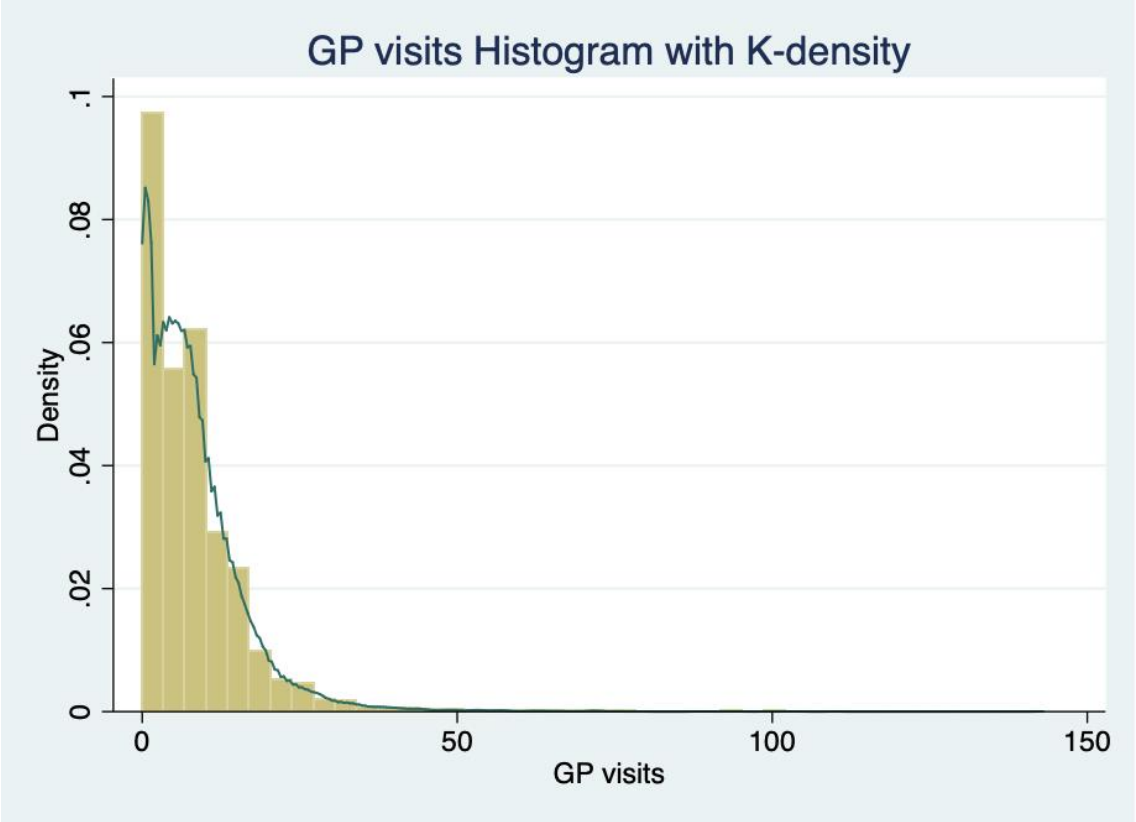


Figure 3-8: Histogram of frequency of GP visits exposure measured as a discrete variable.

As the distribution is not normally distributed, assessment of histograms with transformation was made using the "gladder" code from Stata (Figure 3-9).

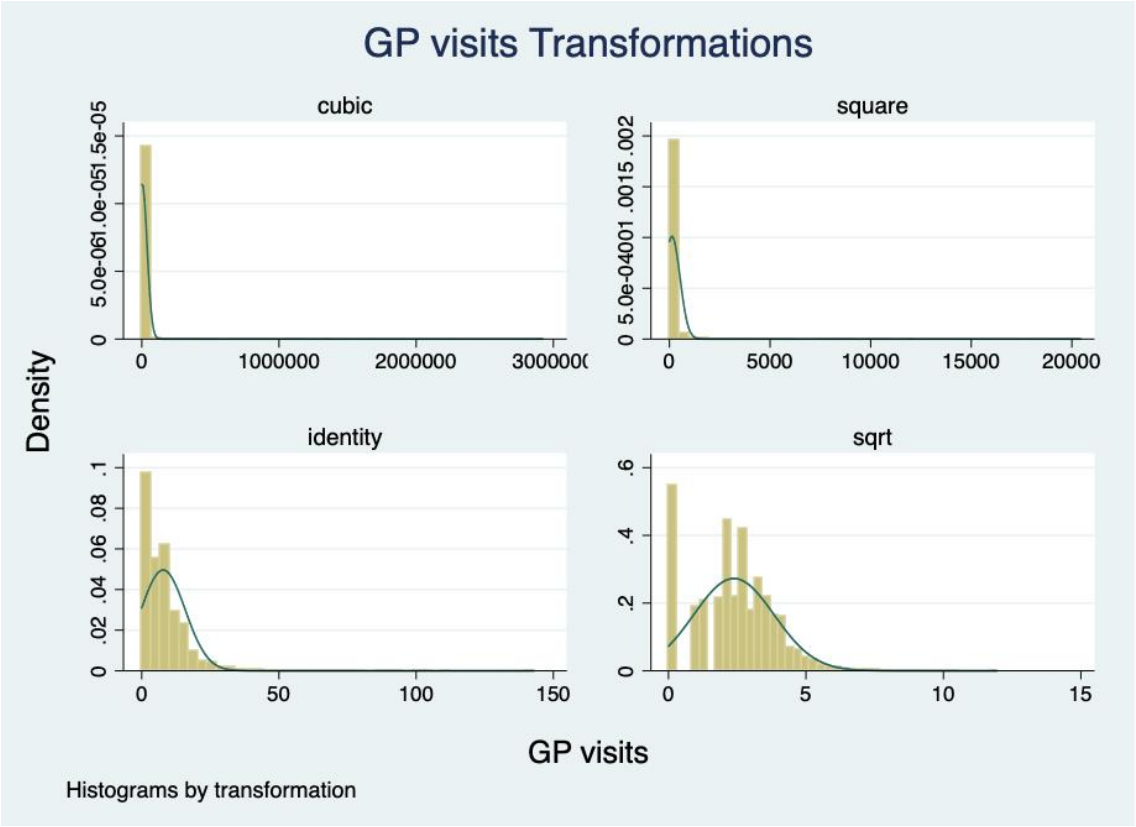


Figure 3-9: Transformations of frequency of GP visits exposure measured as a discrete variable.

There were no transformations that yielded a normal distribution and thus the exposure variable was categorised into the maximum number of groups while ensuring there was sufficient power.

Furthermore, the GP visits variable was analysed using fractional polynomial to assess the distribution of GP visits within the first post-operative year against revision as the

outcome of interest (Figure 3-10). A U-shape associated was noted from the fractional polynomial assessment and the baseline group for the GP visit variable was adjusted accordingly.

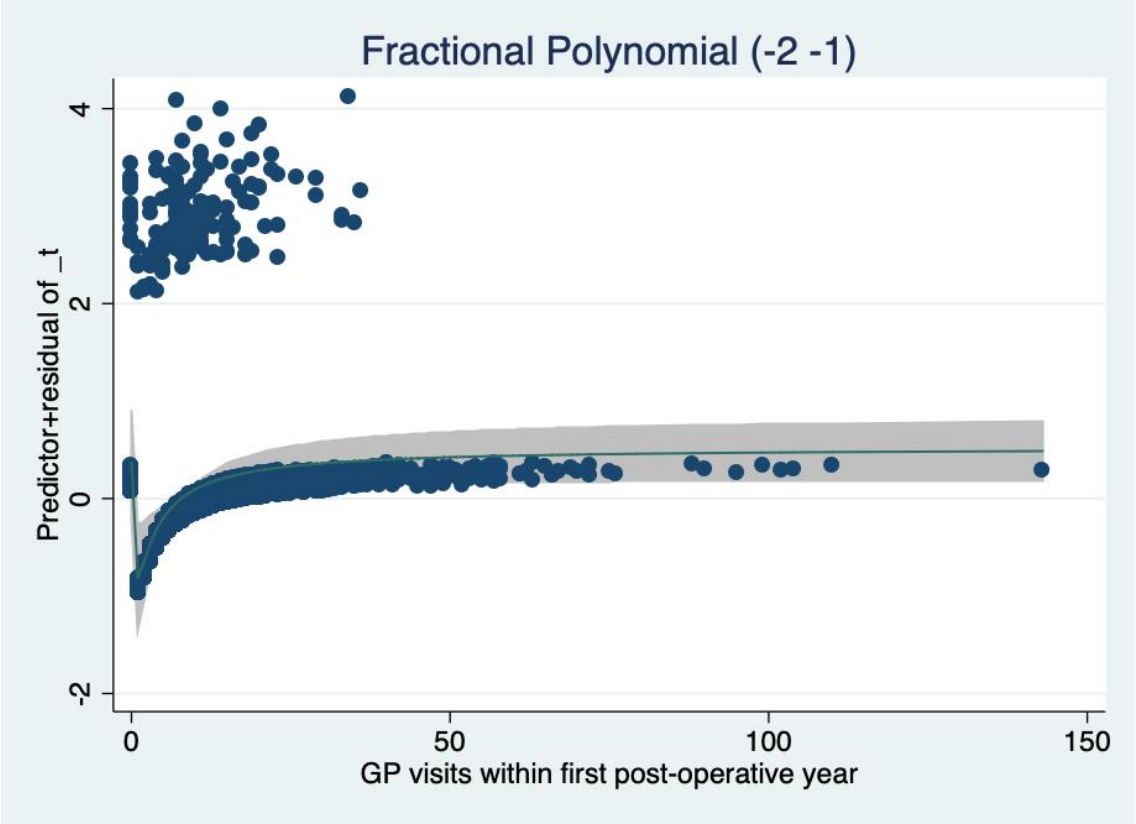


Figure 3-10: Fractional polynomial graph for frequency of GP visits against the outcome of interest (revision).

In Spain, the average annual number of GP visits per patient in 2009 was 4.3[377]. Therefore, GP visits was categorised using a cut-off of four visits per patient per year, with the baseline group as one to four visits due to the U-shape correlation with revision.

The final categorisation of the GP visits variable is Baseline group consist of those patients with one to four visits per patient per year, group one consists of those patients with zero visits per patient per year, and group two consists of those patients with greater than four visits per patient per year.

Referral to the Orthopaedic department

Data was available on the date of referral only for referrals to the Orthopaedic department which was extracted from the SIDIAP database for all eligible patients. Patients with a date of referral after one-year post-operation were not included. Therefore, the variable of referral to the Orthopaedic department must remain binary between patients who were referred and those who were not.

The UK National Health Service (NHS) recommend routine follow-up with the Orthopaedic department for patients who have undergone THA surgery is approximately six to eight weeks post-operatively[345]. The BOA guidelines confirm this with recommending initial follow-up within eight weeks and thereafter yearly for the first five years for novel implants[341]. Unlike the UK, the Spanish healthcare system has a decentralised system[378]. There are no formal guidelines such as those presented by NICE or the BOA for the recommendation of post-operative follow-up in patients undergoing THA. However, Spain has a national healthcare system like the UK, and Spanish literature uses the UK as examples of good practice in THA monitoring[350]. Therefore, UK guidelines were used. Patients referred to the Orthopaedic department within eight weeks following THA surgery were not included within the variable as this could represent those patients that had no routine appointment scheduled as opposed to those patients referred for potential post-operative pain or problems. These

patients were not excluded from the study as they are likely to represent routine post-operative follow-up.

Referral to the Physiotherapy department

Data was available on the date of referral only for referrals to the Physiotherapy department which was extracted from the SIDIAP database for all eligible patients. Patients with a date of referral after one-year post-operation were not included. Therefore, the variable of referral to the Physiotherapy department must remain binary between patients who were referred and those who were not.

Both BOA and NICE recommend significant Physiotherapy team input during the immediate post-operative period until discharge[341,342]. However, no routine post-discharge follow-up is recommended. Furthermore, the majority of patients can return to activities such as driving and working, depending on their job, at six weeks post-operatively[345]. In addition, the Orthopaedic follow-up appointment after the six to eight week post-operative period is at one year, even for novel devices[378]. In light of this, for the majority of patients, it is unlikely Physiotherapy input will be required beyond the six to eight weeks following surgery. Therefore, the referral to the Physiotherapy department variable is treated in the same manner as referrals to the Orthopaedic department, to include only those patients referred after eight weeks and before one year post-operatively.

3.3.8. Confounders

Summary

A priori, variables were identified as potential confounders in relation to failure based on clinical knowledge, literature search, and biological plausibility. Measurements were taken at the time of or prior to surgery as registered by the SIDIAP or RACat databases.

The potential confounding variables identified in the RACat and SIDIAP joint database included: Age, Sex, Socio-economic status, Charlson Co-Morbidity Index, Depression, Atrial Fibrillation, Hypertension, Cardiovascular Disease, Diabetes, Cancer, Indication for Surgery, Previous Fixation, Previous Fracture, Body Mass Index (BMI), Smoking status, Alcohol status.

Introduction

There is significant variability in revision rates for THA devices despite overall rates being low[379]. This variability is due to the potential influence of multiple factors including patient characteristics, prosthesis and hospital factors[379]. Separating pain and disability specific to the hip device rather than from other sources is challenging with one study showing even the Oxford Hip Score can be influenced by co-morbidities[219]. When considering confounding variables, the primary requirement is the existence of an independent relationship between the variable and the outcome of interest[380]. Thus, the primary consideration of confounders is those that could influence revision rate. Despite this, it is important to consider an association with exposure[380], therefore, the potential confounders of revision rate are considered for their association with either

pain or analgesic or healthcare use individually. However, even if a confounder is not known to be associated with an exposure it is possible there is confounding within the dataset and therefore would need to be controlled for in the regression model[381].

Potential confounding variables can be categorised to patient, device, and healthcare related factors.

Patient related factors

Age

Age has been shown to influence revision rates with younger individuals eliciting increased risk for all cause and aseptic loosening[132,382–385]. This has been shown in both cemented and cementless implants[132,238,372,386,387]. While studies primarily show younger patients having a higher risk, older patients at the other end of the spectrum were also at increased risk of revision[388].

Age has been shown to have an association with analgesic use in joint disorders with both NSAID and opioid use found to be decreased in younger patients[389]. NSAID use has been demonstrated to be lower in younger patients following THA surgery, however, older patients have been shown to take more analgesic drugs overall[266]. Another study has confirmed younger patients are less likely to have opioid medication[390]. Furthermore, one study has demonstrated an association with age and subsequent analgesic drug use, for both NSAID and opioid medications, one year following THA surgery[391]. Prolonged opioid use post-operatively was found to be associated with younger age[392]. Finally, age is associated with post-operative pain levels[393].

Younger patients following THA surgery had a higher proportion of neuropathic analgesic use within the first two post-operative years[266]. Gabapentin has been shown to have increased concentrations despite no renal function deficit in older patients[394] and therefore, prescribing in patients over 65 years old may be less favourable. Both Gabapentin and Pregabalin have increased side effects for elderly patients due to poor renal clearance[395]. Older age has been demonstrated as an independent risk factor for neuropathic pain[396]. Furthermore, the most commonly used TCA for neuropathic pain, Amitriptyline, is recommended to be avoided in older patients due to side effects such as sedation and orthostatic hypotension[397].

Increasing age has been associated with increasing healthcare utilisation[398]. Age has been demonstrated to be strongly predictive of high frequency of primary care doctor visits[399]. Older patients are more likely to seek primary care consultations[327], and increased healthcare resource use with those aged over 80 by almost 40 fold[400]. Age has also shown an association with referrals to the Physiotherapy department with increasing rates of referral with increasing age[343]. Older patients are less likely to be referred to the Orthopaedic department even in musculoskeletal specialised primary care clinics[401,402], which has been suggested to be due to clinical uncertainty of the balance for the benefits and risks in this age group[402,403].

Gender

Gender has been demonstrated to influence revision rate with men having higher risks than women[132,238,372,382,383,385,387,388]. One study found the risk was specific to femoral stem revision rather than the acetabular cup[384]. However, the overall association was found mainly for revision due to aseptic loosening compared to all-cause revision[132]. One

large retrospective cohort study using registry data showed women had greater revision risk independent of age[404].

A significant difference for opioid use exists between genders with women having higher use compared to men[405]. In addition, this study confirmed age as an association with use finding older women to have the highest prevalence of opioid use. These findings have been confirmed with women redeeming more analgesic medication of both NSAIDs and opioids, following THA surgery[266,391]. Multiple studies have demonstrated women reporting more pain with lower tolerance or thresholds of pain[406–408]. Women have been demonstrated to experience statistically significant more neuropathic-like pain following joint replacement[409].

Gabapentin and Pregabalin have been shown to be generally used more commonly in women compared to men when treating any neuropathic pain[410]. Both Pregabalin and Gabapentin have a decreased renal clearance in women and may therefore be less favourably used[411]. Tricyclic antidepressants have been shown to have a reduced effect in women with the recommendation for alternative medications to be used[412]. Furthermore, medications used for neuropathic pain has been demonstrated to be higher in women compared to men following THA surgery[266].

Despite more women undergoing THA surgery, there remains inequality in referral patterns perhaps due to a reluctance for operative intervention[413,414]. Furthermore, it has been demonstrated women delay presentation prior to seeking joint replacement[413]. Female gender has been shown to be predictive of high frequency of primary care doctor visit[327,399,415]. This remains the case for musculoskeletal conditions[416]. However, one study has found men suffering with frequent pain had increased rates of visiting their GP[417]. Gender has been shown to be potentially associated with rates of referral to

the Physiotherapy department with women being referred more often than men throughout all age groups[343,418]. Women are associated with a reduced risk of referral to secondary care for those suffering from hip pain[402].

Socio-economic status

Socioeconomic status[419] was shown to influence revision risk with those patients having better social support with more resources at their disposal having a lower risk of revision[388].

Socio-economic status is associated with increased pain at presentation prior to THA surgery,[420] and increased prevalence of pain[421]. The finding of association between lower socio-economic status and higher pain is robust[422]. Despite this, social deprivation was shown to be a significant independent predictor of mean improvement in Oxford hip score following hip arthroplasty when compared to prior scores[420]. However, Robinson et al, showed overall those who were more deprived had poorer outcomes. Socioeconomic factors are associated with significantly worse pain and function after THA surgery[423]. Higher levels of socio-economic status have lower probabilities of analgesic drug prescriptions[424]. In addition, lower socio-economic status predicted higher use of NSAID medication following THA surgery[425]. Opioid use was also found to be significantly higher in those patients of lower socio-economic status[390]. Lower socio-economic status was found to predict prolonged opioid medication use post THA surgery[392]. Neuropathic pain has been demonstrated to be more prevalent in patients of lower socio-economic status[396,426] and therefore, these patients may have a higher rate of atypical analgesic use.

Patients of lower socio-economic status have been shown to be in overall worse health[427] and more likely to suffer from musculoskeletal pain[428]. Furthermore, these patients have been demonstrated to seek consultation with their GP more often compared to patients with higher socio-economic status[415,417,429,430]. Socio-economic status has been shown to be associated with healthcare resource utilisation[400,431], with those of lower status having higher usage[398,432]. Furthermore, those of lower socio-economic status are associated with lower risk of referral to secondary care for hip related pain[402].

Co-morbidity

Multiple comorbidities including renal disease, chronic urinary tract infection, and congestive heart failure have been shown to be independent risk factors[433]. Diabetes has been shown to increase revision risk in THA[326,434] which may be due to deep infection[435]. This increased risk of infection post-operatively could be caused by peripheral vascular disease and cerebrovascular disease[435]. Comorbidity scores have been shown to be associated with revision rate in THA[436]. The Charlson comorbidity index score is one of the most investigated and validated[132,370,384,387]. The Charlson score includes a variety of comorbidities such as peripheral arterial disease, cerebrovascular disease, diabetes and cancer[436]. An abbreviated version of the Charlson score adapted by the Royal College of Surgeons (RCS) has been developed[437]. The RCS Charlson score was associated with a worse outcome for late revision rates[436]. A large national registry based study found cardiovascular disease, hypertension, and cancer were independent risk factors for revision rate in hip arthroplasty surgery[434]. In this study, cardiovascular disease included coronary heart disease, atrial fibrillation, and heart failure.

Hypertension was defined in the absence of concomitant cardiovascular disease. While this study found no significant independent risk with hypertension and atrial fibrillation, other studies have demonstrated their potential to influence early revision rate independently[326] due to infection and are thus important confounders[438–441].

Increased number of co-morbidities, as measured by the abbreviated Charlson co-morbidity score, has been shown to be associated with increased use of analgesic medications such as NSAID and opioid drugs in the first year post THA surgery[391].

A higher Charlson co-morbidity score has been demonstrated to increase the risk of neuropathic pain for musculoskeletal conditions[442]. Furthermore, it has been shown to increase the risk of neuropathic pain following THA surgery[443].

Charlson co-morbidity index has been shown to be associated with frequency of visits to a primary care doctor[399]. In addition, the Charlson score has been demonstrated to predict resource utilisation[444]. Patients with a higher Charlson score have increased risk of complications and prolonged length of stay following THA surgery[445]. Therefore, GPs may delay referral in preference of primary care management to avoid potential complications. This is potentially supported as patients with multiple comorbidities suffering from hip pain have been shown to have a reduced risk of referral to secondary care[402]. In addition, patients with multiple comorbidities have an increased risk of higher healthcare utilisation in general[398] including higher frequency of primary care visits[446–448]. These patients are also more likely to be referred to secondary care for consideration of joint replacement[449].

NSAID use was shown to be less likely in those patients with hip pain who in addition, had cardiovascular disease[450]. Furthermore, NSAID use following THA surgery is significantly less in patients with cardiovascular disease[391]. NSAID use is a significant

risk factor for atrial fibrillation and clinicians are advised to avoid its use in patients with or at risk of developing this condition[451,452]. One study suggested that an association of BMI with hypertension could be a reason for the increased level of pain post-operatively[453].

Ultimately, NSAID prescribing must be considered with caution due to the potential for cardiovascular risks and therefore prescribing habits would be influenced (either not prescribed, reduced dose or shortened length of prescription) to avoid adverse events or drug interaction[260]. Despite these concerns, NSAID use is still more common as there are risks with opioid medications and thus NSAID may be more favourably prescribed by clinicians in patients with cardiovascular co-morbidities[260,454]. This would explain lower than expected use of opioids following THA surgery[266].

Gabapentin and Pregabalin have been shown to increase the risk of Atrial Fibrillation even in the absence of previous cardiovascular disease[455] and would thus potentially be avoided in patients with any cardiovascular, or atrial fibrillation history. The British National Formulary (BNF) advises both Pregabalin[456] and Amitriptyline[457] should be prescribed with caution in patients suffering with cardiovascular disease and thus could be prescribed less favourably in these patients. There are cardiac toxicity concerns for patients taking TCAs and thus they are recommended to be prescribed with caution in those patients with cardiovascular disease[458].

Patients who have hypertension or heart disease have been shown to be associated with higher healthcare resource utilisation[398,400]. Patients with hypertension have been demonstrated to have higher primary consultation rates[459]. NICE guidelines recommend routine review of patients suffering with hypertension[460] which may increase primary care consultation rates. During the maintenance phase of treatment

for atrial fibrillation, 80% of patients are seen by their primary care physician rather than a nurse or specialist clinic[461]. These patients are, therefore, potentially more likely to have increased consultation rates compared to the general population.

A pre-operative diagnosis of diabetes has been shown to be an independent risk factor for persistent pain following THA surgery[462]. Diabetes has also been demonstrated to be associated with slower recovery post-operatively in regard to pain[463]. Interestingly, HbA1c score was not associated with subsequent joint pain[462]. NSAID medications have potential adverse events due to interactions with medication used in treating diabetes and may therefore be less readily prescribed[464]. In addition, there is a recommendation to avoid post-operative use of opioid medication when possible[465]. However, prolonged opioid use following THA surgery was associated with pre-operative diabetes[392]. It is important to note that obesity (BMI), hypertension and diabetes form part of the metabolic syndrome which is common in patients undergoing joint replacement[466]. However, the literature search performed here have shown each as potential independent risk factors of pain or analgesic use post-operatively.

Approximately 25% of patients who have developed diabetic neuropathy also have neuropathic pain[467]. Diabetes is associated with neuropathic pain for which Amitriptyline, Gabapentin and Pregabalin medication are commonly used treatments[376]. Amitriptyline has been shown to be an independent risk factor for diabetes[468] and may be avoided in these patients. The BNF advises Amitriptyline should be prescribed with caution in patients suffering with diabetes and thus could be prescribed less favourably in these patients[457]. The BNF also advises caution in prescribing Gabapentin in patients with diabetes[469].

Patients who suffer with diabetes are associated with increased healthcare resource utilisation[398,400]. Furthermore, NICE guidance recommend regular routine follow-up for patients suffering with diabetes[470] and therefore are potentially more likely to have higher rates of primary care consultations. Patients with diabetes have significant negative impacts following Orthopaedic surgery[471]. Therefore, GPs may be more inclined to manage these patients in primary care or refer for Physiotherapy as an alternative conservative option.

Depression

Depression and psychoses have separately been shown to be independent risk factors of revision post THA surgery[433,434]. Further studies have confirmed depression and anxiety as independent risk factors of revision[326]. Depression has been shown to predict prolonged pain and worse PROM scores and would thus be an important confounder for the exposures investigated here[472].

Psychological disorders such as depression is associated with prevalence of pain in musculoskeletal disorders[421]. Pre-operative anxiety has been shown to be a significant predictor of post-operative pain[393]. Mental health disorders such as depression prior to surgery has been associated with increased post-operative opioid drug use[390]. Patients undergoing THA surgery who have depression and/or anxiety pre-operatively have been found to have poor long-term pain post-operatively[473].

A significant proportion of patients have depression and anxiety prior to THA surgery with a prevalence between 5% and 15%[259,474]. Both depression and anxiety have been shown to be important predictive factors for persistent neuropathic pain[475].

Furthermore, mental health issues prior to THA surgery are associated with post-operative neuropathic pain[476,477].

Patients suffering from mental health disorders have a higher risk of increased healthcare utilisation[398]. Anxiety and depression have been demonstrated to be independently associated with higher healthcare utilisation[400,478]. In addition, patients suffering with anxiety are associated with increased GP visits specifically[417,479]. Depression and anxiety have been demonstrated to be associated with poor adherence to Physiotherapy treatment[480] and these patients may be less readily willing to accept referrals.

Body Mass Index (BMI)

Weight and obesity have been shown to influence revision rate[238,326,386,388], although it is unclear whether there is interaction with other comorbidities. Height has been shown to also be a potential risk factor[383]. Body Mass Index, combining both weight and height, has been shown to independently increase revision rate[382,384,385,388,481–484] and influence early mortality rates in post THA surgery[485].

Non-opioid analgesic medication such as NSAIDs have been shown to be more commonly used in obese persons[450]. Increased BMI has been shown to predict higher use of both NSAID and opioid drugs in the first post-operative year following THA surgery[391]. In addition, higher BMI pre-operatively was associated with poorer levels of pain control and functional outcomes following THA surgery[423,473].

BMI has been shown to be an independent risk factor for neuropathic pain in musculoskeletal conditions of the hip[289]. Thus, patients with a higher BMI may require increased rates of atypical analgesic use.

An increased body mass index, for those categorised as very or severely obese, have been demonstrated to have a statistically significant increased risk of higher healthcare utilisation[400]. In addition to those who are obese, underweight patients have been associated with increased healthcare utilisation[398]. Despite current evidence being contradictory, obesity may lead to increased rate of complications following Orthopaedic surgery[383]. Therefore, the likelihood of referral for these patients may be reduced[402].

Cancer

Opioid use for treating joint pain was found to be more likely in those patients who had a history of cancer[450]. Furthermore, opioid medication is preferred over NSAIDs when treating prolonged pain in patients with previous cancer[486]. These factors could influence prescribing practices in patients following THA surgery who have a history of cancer.

Neuropathic pain is common and often debilitating in patients with cancer[487] occurring in up to 90% of patients following chemotherapy[488]. Thus, the presence of cancer can increase the potential requirement for neuropathic analgesic use. Amitriptyline, Gabapentin and Pregabalin are all recommended as adjuvant medication in the management of cancer related neuropathic pain[489]. However, the evidence for the

effectiveness of Amitriptyline for neuropathic pain in patients with cancer is minimal[490].

Patients with current or previous cancer has been shown to be associated with increased healthcare resource utilisation[398,400]. One study has demonstrated a statistically significant increase in risk of requiring THA surgery following a diagnosis of cancer which may be a direct cause from treatment received[491]. Furthermore, patients with a history of cancer are more at risk of developing complications following Orthopaedic surgery[492], but this may not be the case for all cancers[493]. These factors could influence referral rates for patients with previous cancer to both the Orthopaedic department and to Physiotherapy, which would provide an alternative non-operative option.

Indication for surgery

Underlying diagnosis as the indication for THA surgery has been shown to influence survivorship in both the acetabular and femoral components[372,382,387,404,494]. Further studies demonstrate particular diseases, such as Rheumatoid Arthritis or Avascular Necrosis, have increased revision rates[132,326,433].

NSAID use is clearly indicated for inflammatory conditions such as Rheumatoid Arthritis (RA) but less rigorously proven for osteoarthritis[495]. It is conceivable that clinicians may increase dose of NSAIDs for patients, but move to a different agent, such as Tramadol, or suggest alternative methods, such as weight loss[496], more readily for patients suffering from osteoarthritis.

There have been few studies evaluating the use of Amitriptyline in musculoskeletal disorders but have shown significant improvement in neuropathic pain when compared to Gabapentin or Pregabalin use and may thus be favoured in patients with Osteoarthritis[497]. One potential side effect from TCAs is a dry mouth which can be of specific concern for patients suffering from RA and secondary Sjögren's syndrome[498]. Therefore, it is possible these drugs are favoured less in patients in these patients.

Patients suffering from arthritis are more likely to see their GP and those suffering from Rheumatoid Arthritis have a higher rate of referral to a specialist[415,416]. Early referral is recommended for patients with Rheumatoid Arthritis[499,500], thus GPs are likely to seek specialist help in managing any patients with post-operative issues. However, outcomes following revision THA is significantly worse in patients suffering from Rheumatoid Arthritis[501], thus both patients and GPs may be more inclined to delay surgery opting for conservative management such as Physiotherapy.

Previous fixation

Previous fixation[385,483] was shown to be a risk factor where the second hip replacement on the contralateral side, had increased risk of revision compared to the first especially if operated a few years later[502].

In patients with contralateral hip osteoarthritis with or without previous fixation a significant increase in the levels of pain following THA surgery has been found[423,503,504]. Therefore, it is possible for such patients to be expectant of analgesic medications due to prior experience.

Although limited, there is evidence in the use of antidepressants, such as Amitriptyline, in the post-operative period to reduce acute or chronic post-surgical pain. Patients who have had such medication previously may influence subsequent use following another operation[505]. Furthermore, while rare, previous THA surgery is a risk factor for nerve injury and subsequent neuropathic pain[506].

Patients who perceive that there may be more serious consequences due to their condition are more likely to be referred to a specialist[330]. Therefore, those who have had previous fixations and perceive their condition to be severe may seek referral earlier or more readily. Patients who have had discussions regarding joint replacement are more likely to be referred for consideration of surgery[507]. Furthermore, GPs are uncertain regarding which patients are appropriate for consideration of THA surgery[508]. Therefore, GPs may be more inclined to refer patients early when faced with the complexity of post-operative issues. Patients who are affected by Osteoarthritis in multiple joints are at higher risk of being referred for joint replacement[449]. Furthermore, patients with multiple symptomatic joints are more likely to suffer following joint replacements[509] and are therefore less likely to seek surgical options opting for non-operative treatments such as Physiotherapy[449].

Previous fracture

Previous fracture were associated with mixed risks of revision but shown to influence rates[382,387,510].

Patients who suffer from arthritis are more likely to experience a fracture when using opioids compared to using NSAID[511]. Therefore, patients with a previous fracture may be more likely to be prescribed NSAID compared to tramadol following THA surgery.

There is a high prevalence of neuropathic pain in patients with previous fracture and operative treatment[512]. Persistent pain after one year post-operatively is common following previous fractures[513]. Furthermore, anticonvulsant medication, such as Gabapentin and Pregabalin, have been shown to increase the risk of non-traumatic fractures[514]. Therefore, there may be avoidance to prescribe these medications in patients with previous fractures.

The referral rate for those patients with trauma and related conditions is lower than for arthritis[416]. Furthermore, those with trauma had an increased rate of primary care visits due to "other sprains and strains", and were more likely to be referred to a specialist[416]. Patients who suffer with Osteoporosis are likely to have had fractures[515] especially in the cohort who undergo joint replacement. Furthermore, GPs are uncertain regarding which patients are appropriate for consideration of THA surgery[508]. Therefore, GPs may be more inclined to refer patients early when faced with the complexity of post-operative issues in those who have had previous fractures and may have Osteoporosis. In addition, patients who have had fractures are at increased risk of having early post-operative periprosthetic fractures[516] which would require referral to the Orthopaedic department.

Smoking status

Smoking has over a four times risk of revision in THA[517] which may be due to increased risk of infection leading to early revision[518], but the effect may be less so in long-term revision[519].

Smoking has been shown to be associated with higher analgesic use[519]. In addition, there is an association between patients who smoke and increased pain perception[520]. Patients who smoke are associated with increased reporting of musculoskeletal pain and subsequently the increased prevalence of pain[421].

One study has demonstrated patients with neuropathic pain are more likely to be smokers[396]. Furthermore, smoking has been shown to increase the risk of developing neuropathic pain[521] and thus, potentially increase the likelihood of these patients being prescribed neuropathic analgesic. There has been preliminary research suggesting the potential use of both Gabapentin and Pregabalin in helping patients with smoking abstinence[522,523] and thus may be used favourably by clinicians.

Smoking has been shown to be associated with healthcare utilisation[398,524], with any level of smoking having an increased risk of higher usage[400]. Furthermore, due to the potential for increased risk of post-operative complications for smokers undergoing Orthopaedic surgery[525,526], referral is less likely in this subset of patients[402].

Alcohol status

Alcohol use was also associated with increased revision[265] as well as other post-operative complications[527].

For patients with higher levels of alcohol use following surgery, non-opioid medications are recommended to be given as a priority due to a "more favourable risk profile" with additional benefits of improving anxiety and depression for this subset of patients[528]. The "co-use of opioids and alcohol" has been demonstrated to produce worse outcomes in the treatment of pain[529]. These factors could influence prescribing practices following THA surgery.

Pregabalin and Gabapentin have been found to have benefits beyond simple analgesic effects in patients with higher alcohol consumption following surgery, including reducing alcohol cravings and prolonging abstinence from alcohol[528]. If NSAID were deemed potentially unsafe for a patient medications such as Pregabalin could be a suitable alternative especially if a neuropathic component was present[528]. Recommendations goes further to suggest non-opioid analgesic medications should be given priority in these patients[528].

Of note, use of other analgesics such as NSAID and opioids were not shown to be independent risk factors for neuropathic pain[289].

Alcohol status has been shown to be associated with healthcare utilisation and for those who are categorised as heavy drinkers having reduced usage[400]. In a similar manner to smoking, patients with significant alcohol use have a higher risk of complications[526] and would therefore be more likely to be referred to the

Physiotherapy department or managed in primary care longer for Orthopaedic issues compared to the Orthopaedic department.

Device related factors

The recent issues with MoM bearings has been discussed previously and highlights the potential for the type of bearings to influence revision rates[46]. This is confirmed in other studies[482].

Femoral head size has been shown to influence the risk of revision although a stronger association may exist for revisions due to dislocation[132,404,530]. The size of the femoral stem was also shown as independent risk factor[482].

Fixation type (cemented or cementless) has been shown to influence revision rate for both acetabular and femoral componenets[483,494,531,532]. Hybrid fixation technique was shown to potentially have an influence on revision risk when compared to cemented or uncemented fixation[532,533].

Healthcare related factors

Hospitals performing less than 200 THA operations annually had a 34% increased risk of early aseptic revision compared to hospitals performing more than 400 operations annually[224]. The finding of hospitals performing more THA procedures per year with lower revision rates is not unique[534,535].

Surgeons performing lower volumes of THA surgery per year were also found to have increased risk of revision[404,534,536,537].

Longer operating times were shown to significantly increase revision rates but a short operating time could increase risk of revision for cemented implants[538].

Conclusion

Despite there being numerous confounding factors that may need to be considered, the variables used within this thesis were limited to availability and analytical potential. Furthermore, for each category of confounder included in the regression model, there must be at least two or more outcome events in each category of the confounding variable[539,540].

BMI, smoking status and alcohol status variables have significant missing data within this cohort. Therefore, multiple imputation by chained equations was required to impute missing data and perform analysis on a complete dataset.

3.3.9. Statistical analysis

Summary

Descriptive statistics was used to describe the cohort sample and exposures. For the three confounding variables with substantial proportions of missing data (up to 40%), multiple imputation by chained equations was applied to ensure analysis can be performed on a complete cohort to reduce the risk of selection bias and increase precision from the studies being otherwise potentially underpowered.

All variables included in subsequent regression analysis (exposures, confounders, and outcome) and any identified auxiliary variables have been included within the multiple imputation model. 40 imputations were performed using the inbuilt Stata command. Evaluation of the multiple imputation was performed by comparing pre- and post-multiple imputation univariate modelling and Monte Carlo Errors. Subsequent main analysis was performed on individual datasets and combined in conjunction with Rubin rules using in built Stata commands.

Survival analysis was used to model the association between exposures and implant survival. Lifetables were constructed and reported numbers at risk and revised in each year of follow-up. Kaplan-Meier estimates of the cumulative incidence of revision surgery stratified by exposure in the first post-operative year were plotted. Fine and Gray (F&G) regression, accounting for a competing risk with death[541], was used to estimate sub-hazard ratios (SHR) and 95% confidence intervals according to exposure. Multivariable F&G models were used to adjust for pre-identified confounders as previously discussed. Sensitivity analysis was performed for a priori variables including analysis in patients with diagnosis of osteoarthritis only, type of bearing and

fixation type (if appropriate). All statistical analysis including multiple imputation was performed using Stata 15 (StataCorp. 2017)[542] statistical software.

A more detailed explanation and argument for multiple imputation methods, and for the use of survival analysis and F&G modelling are found in Appendix 2 and 3 respectively.

Post-estimation evaluation of F&G modelling

The studies here are primarily investigating an association between new exposure variables that could potentially be surrogate markers for risk of subsequent revision. There are post-estimation tests that are needed to evaluate the performance of the regression model.

While these exposures are not diagnostic tests in themselves, it is possible to assess how well the exposure can differentiate which implants are more likely to be revised. One of the simplest methods to quantify the exposure's diagnostic ability is to assess the proportions of positive (presence of revision) or negative (absence of revision) results. There are two measures commonly used[543]. The first is known as 'sensitivity' which represents the proportion of correctly identified positives by the test. The second is known as the specificity which represents the proportion of correctly identified negatives by the test.

However, sensitivity and specificity do not indicate the probability of a test giving the correct diagnosis. Therefore, predictive values can be used[544]. These include 'positive predictive value' (PPV) and 'negative predictive value' (NPV). Here, the PPV is the proportion of patients who have 'tested positive' for revision and are correctly identified. The NPV is the proportion of patients who have tested negative for revision surgery and are correctly identified.

The majority of diagnostic tools are quantitative and the extent of which the results differ from patients who do or do not have revision can be investigated[545]. One method of doing so is by plotting the receiver operating characteristic (ROC) curve[545]. The ROC curve is created by plotting sensitivity against 1-specificity. A global assessment of this curve can be generated by calculating the area under the ROC curve. This measure represents the probability that a randomly selected patient who has undergone revision surgery has a higher value of the exposure than a randomly selected patient who has no revision operation. A probability of 0.5 is considered an "uninformative test"[545]. A value of 1 is perfect discrimination[546].

As with other survival models[547,548], the main assumption of F&G regression is that the sub-hazard remains proportional over time[549,550]. If at a specified time one device has a twofold risk of revision compared to another device, this risk will remain twofold at other time intervals during the study period. There were two methods used to investigate this assumption. Firstly, graphs of the cumulative incidence function were generated comparing exposed to non-exposed patients. Crossing of the survival curves indicate the proportional sub-hazards assumption has been violated[547]. Secondly, Stata

statistical software has an inbuilt command ('tvc' option), to test the sub-proportional hazards assumption following F&G analysis[551] which was used here.

Sensitivity analyses

The main indication for total hip arthroplasty is osteoarthritis[130,238,240] and new implants are primarily used in patients for this indication[552]. Therefore, a sensitivity analysis was performed investigating the association of exposures with subsequent revision rate for patients undergoing primary THA for osteoarthritis only.

One of the main reasons for the investigations of potential surrogate markers of primary THA revision is the recent metal-on-metal disaster[46]. Therefore, a sensitivity analysis stratifying results by bearing type was performed to investigate whether differences between MoM and other bearings could be identified.

The distribution of the Charlson score within this cohort was 44.6%, 5.7% and 50.7% for scores of zero, one to two, and more than two respectively. However, this is considerably different to published distribution of the Charlson score within a Spanish THA cohort showing 78.5%, 20.4% and 1.1% for scores of zero, one to two, and more than two respectively[553]. Therefore, a sensitivity analysis was performed investigating the association of exposures with subsequent revision rate excluding the Charlson score as a confounder.

Sample size

While observational studies using registry data do not require recruitment of patients, a sample size calculation a priori is required to ascertain the risk of a type II error[554] of any subsequent analysis.

Calculating the sample size for survival analysis with a binary outcome measure has unique challenges compared to other regression analysis when faced with censoring or violation of the proportional hazard assumption[555]. Therefore, a Stata package called "ART" was used that has been designed to address issues of power calculation in survival analysis[555].

As part of the "ART" package calculation, an estimated effect size is required. The recent MoM devices offer an insight to failing implants. The latest NJR report stated "For every 100 MoM hip-resurfacing procedures, we estimate that there would be 7.8 excess revisions by ten years, and similarly for every 100 stemmed MoM THR procedures that there would be 15.9…"[556]. The maximum revision rate for a proven THA device is 5%[108]. Therefore, a MoM device would have a revision rate of 12.8% versus a proven device of 5% giving an estimated minimum effect size of approximately 2.5. Recent studies have shown the ASR hip device has a 3.64 hazard ratio at 2-3 year mark which corresponds to the median follow-up time in the cohort here[557]. Other studies have confirmed an approximately three-fold hazard ratio of MoM devices compared to standard bearings and the ASR device has a higher hazard ratio of six times more likely to be revisied[558].

Another study has shown the revision rate at three years post-operatively of a standard THA device is 2.1% versus MoM THAs at 4.0% and MoM Hip resurfacing devices at 4.5%[559]. These short-term revision rates were similar in an alternative study[560].

Using these figures would provide an estimated effect size of 1.9 for MoM THA devices and 2.1 for MoM Hip resurfacing devices.

Therefore, the lowest effect size of 1.9 was chosen to calculate sample size. Using a log rank test for equality of survival curves, with 80% power at 5% level of significance (alpha) (p-value of 0.05), assuming equal size groups, and a failure rate in control group of 5%, with a HR of 1.9 compared to MoM group, with 20% expected loss to follow up at 10-years, gives a sample size of 1,264 patients overall, or 632 patients in each group. The number of subjects available in our dataset exceeded this required sample size, therefore providing sufficient statistical power (>80%) for the proposed analyses.

3.4. Results

3.4.1. Confounders

Continuous variables

Age

Age was a continuous variable calculated from the date of birth of the patient. A histogram of the variable showed a slightly negatively skewed distribution (Figure 3-11).

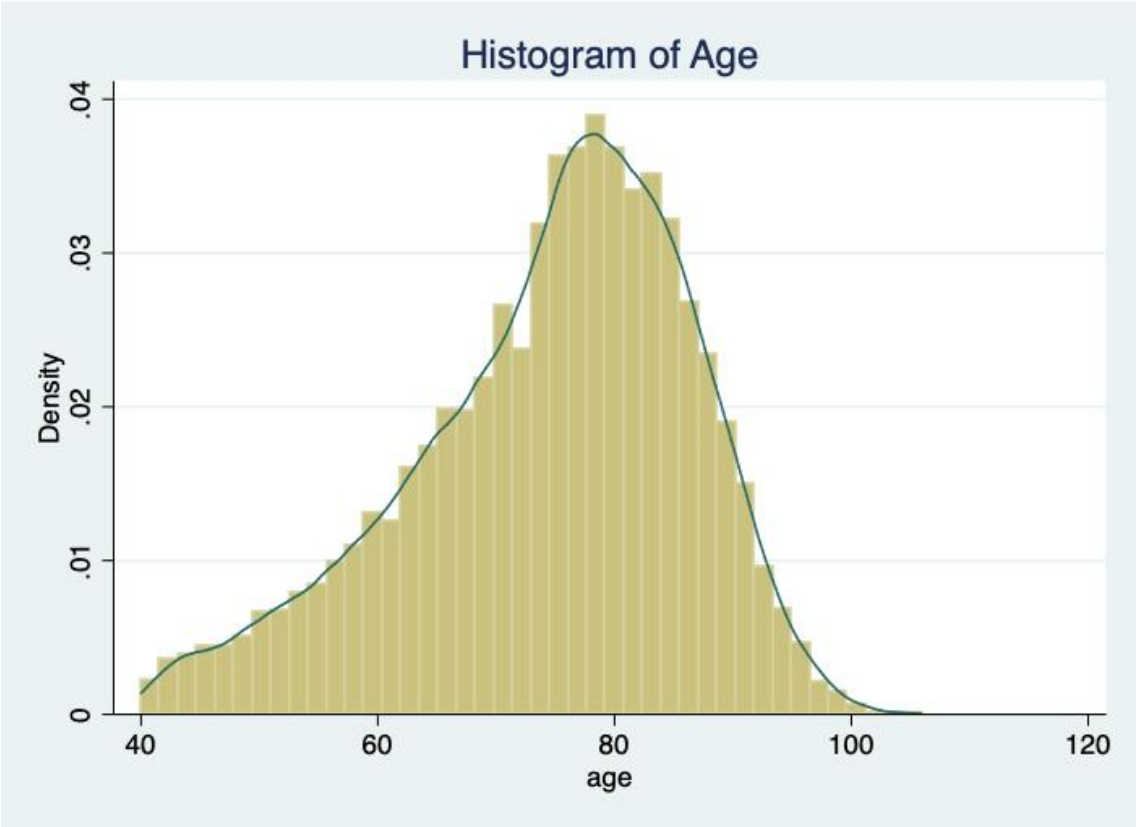


Figure 3-11: Histogram of age variable.

No transformations identified a normal distribution (Figure 3-12). Therefore, the age variable was categorised.

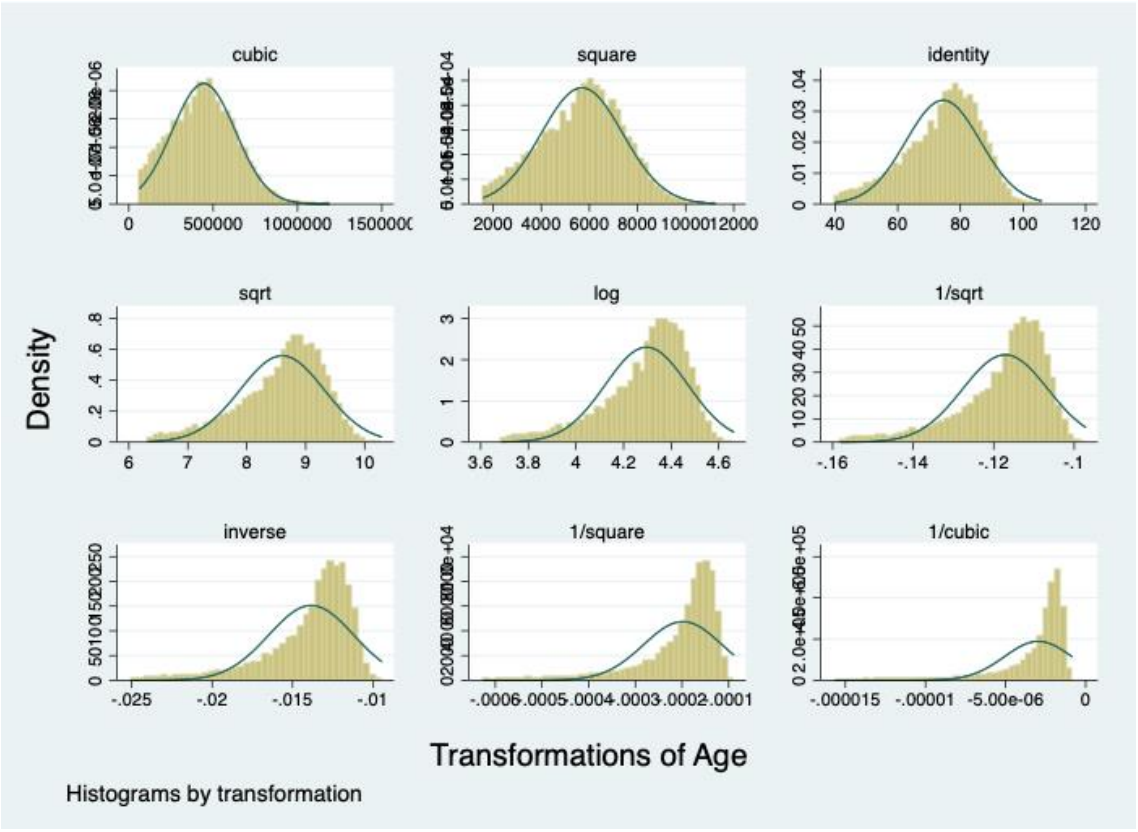


Figure 3-12: Transformations of age variable.

Categories for age were in groups of 10 years starting from age 60. The lowest category was from aged 40 to 60. Most revision operations occurs in patients aged 60-80[242]. The highest category was those aged over 80[533]. Age categories have a greater clinical relevance for comparison[404].

Body mass index (BMI)

BMI was a continuous variable, and a histogram showed an acceptable normal distribution (Figure 3-13). This variable remained a continuous one.

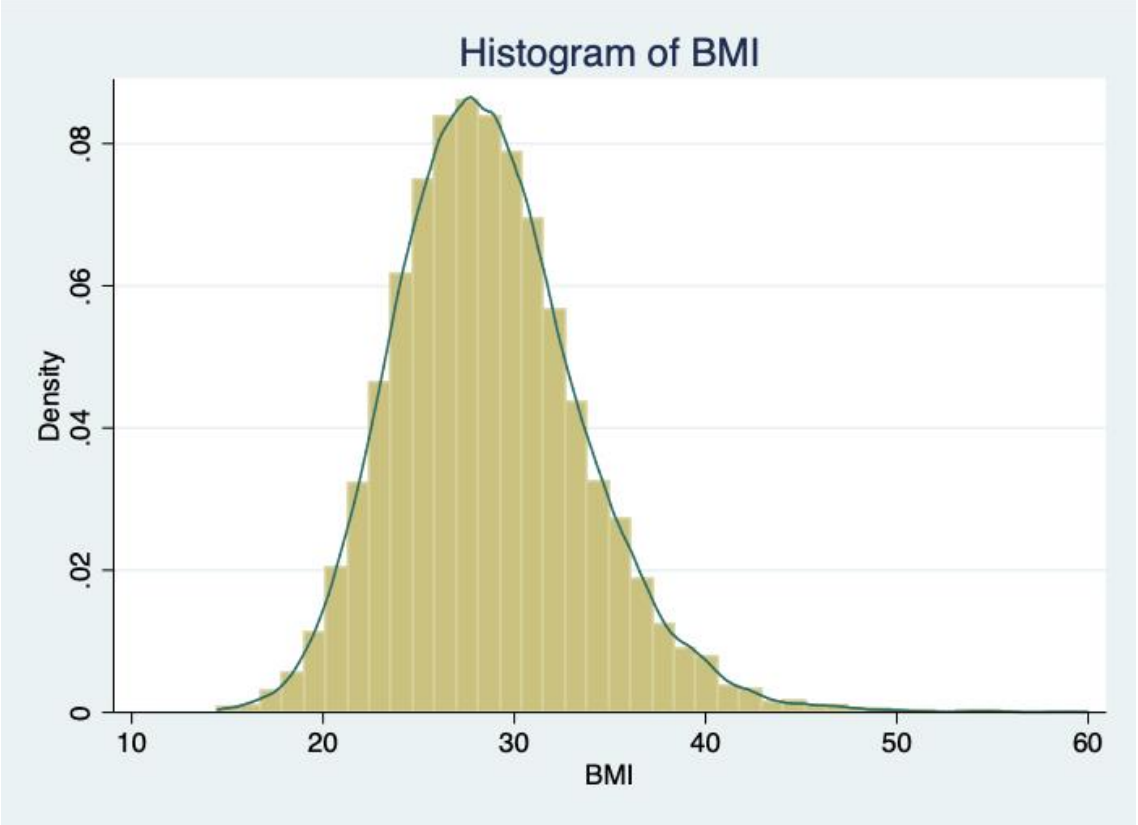


Figure 3-13: Histogram of body mass index variable.

Socio-economic status

The socio-economic status is calculated using the Deprivation Index (the MEDEA project)[561]. Although this variable is continuous it is commonly converted to a categorical variable for investigations and interpretation[562]. In this cohort, the missing

values were not missing but of those patients in a rural environment. Thus, a sixth category to include rural patients only was created.

Categorical variables

The following variables were in categorical form: Sex, Charlson Co-Morbidity Index, Depression, Atrial Fibrillation, Hypertension, Cardiovascular Disease, Diabetes, Cancer, Indication for Surgery, Previous Fixation, Previous Fracture, Smoking status, Alcohol status.

Comparison with revision

All categorical variables were investigated for the number of outcome events per category as too few can render regression modelling invalid and a minimum number of two events is required[539,540] (Table 3-1).

Table 3-1: Comparison of confounders with outcome variable (revision).

Characteristic	Category	Revision No	Revision Yes
BMI	Available	10,816	108
	Missing	5,459	50
Age	40 to <60	2,123	35
	60 to <70	2,884	44
	70 to <80	5,345	54
	80 and older	5,923	25
Gender	Male	6,397	70
	Female	9,878	88
Alcohol status	No	7,198	54
	Yes	2,396	14
	Ex	239	1
	Missing	6,442	89
Smoking status	Never	9,097	91

	Current	1,267	18
	Ex	1,819	17
	Missing	4,092	32
Charlson Score	0	7,259	78
	1	757	10
	2	611	8
	3	239	1
	4	725	4
	5	637	1
	6	1,480	9
	7	4,567	47
Socio-economic status	1	2,731	43
	2	2,748	25
	3	2,748	20
	4	2,744	26
	5	2,745	25
	Rural	2,559	19
Transient Ischaemic Attack (TIA)	No	15,937	155
	Yes	338	3
Ischaemic Heart Disease (Myocardial Infarction or Angina)	No	15,607	154
	Yes	668	4
Peripheral Arterial Disease	No	15,969	157
	Yes	306	1
Atrial Fibrillation	No	15,084	154
	Yes	1,191	4
Hypertension	No	7,303	88
	Yes	8,972	70
Diabetes	No	13,310	132
	Yes	2,965	26
Depression/Anxiety	No	13,096	127
	Yes	3,179	31
Cancer	No	15,583	154
	Yes	692	4
Indication for Primary THA	Fracture	5,157	43
	Osteoarthritis	9,918	99
	Rheumatoid Arthritis	266	5
	Other Arthropathy	199	0
	Congenital Bone Disease	426	9
	Acquired Bone Disease	73	0
	Neoplasm	8	0
	Osteomyelitis	2	0
	Mechanical Complications	132	1
	Other	94	1
Previous Fixation	No	15,496	148
	Yes	779	10
Previous Fracture	No	13,805	140
	Yes	2,470	18

Alcohol status had an 'ex-user' category with only one revision event which is below the threshold. Therefore, this was combined with current drinkers for a new category of current or ex-drinkers. Smoking status had revision events over the threshold required, however, as multiple imputation is used to account for missing data within this variable, it was also categorised into a binary variable[563].

Charlson score had categories with too few or no outcome events. This variable was abbreviated in line with the Charlson score that has been adapted for the surgical setting by the Royal College of Surgeons[437]. Other independent risk factors for revision, as identified earlier, are included within the Charlson score such as diabetes or myocardial infarction. However, as the Charlson score is now abbreviated and these have been shown to be independent risk factors, these individual variables are included as confounders for the final analysis.

Transient ischaemic attack, ischaemic heart disease and peripheral arterial disease variables can be considered as cardiovascular disease[564,565] and were combined as such due to low outcome events per variable in the peripheral arterial disease variable.

Indications for primary THA had multiple categories with too few outcome events per variable. This variable was categorised as patients with fracture, osteoarthritis, or other indications for the final analysis.

Table 3-2 shows variables that have been further categorised following assessment against the outcome of interest.

Table 3-2: Comparison of adjusted confounders with outcome variable (revision).

Characteristic	Category	Revision No	Revision Yes
Alcohol status	No	7,198	54
	Yes/Ex	2,635	15
Smoking status	Never	9,097	91
	Current/Ex	3,086	35
Charlson Score	0	7,259	78
	1	757	10
	≥2	8,259	70
Cardiovascular Disease	No	15,057	150
	Yes	1,218	8
Indication for Primary THA	Fracture	5,157	43
	Osteoarthritis	9,918	99
	Other	1,200	16

Device and healthcare risk factors

Risk factors for revision such as hospital or surgeon's volume of operations per year, the type of device or fixation or bearing where found to have no potential association with exposures and therefore not included as confounders. However, subsequent sensitivity analysis was conducted using device related risk factors and the distribution of these variables are found in Table 3-3.

Table 3-3: Characteristics of device related risk factors.

Characteristic	Category	Observation N (%)
Bearing Type	Metal-on-Metal	534 (3.2)
	Metal-on-Poly	7,160 (43.6)
	Ceramic-on-Ceramic	606 (3.7)
	Ceramic-on-Poly	1,136 (6.9)
	Missing	6,997 (42.6)
Fixation Type	Cemented	3,522 (21.4)
	Cementless	7,821 (47.6)
	Hybrid	2,383 (14.5)
	Missing	2,707 (16.5)
Prosthesis Type	THR	8,862 (53.9)
	Resurfacing	244 (1.5)
	Other	4,629 (28.2)
	Missing	2,698 (16.4)

3.4.2. Multiple imputation

Multiple imputation results for the atypical analgesic cohort can be found in Appendix 4.

Baseline characteristics

Table 3-4 shows the availability of the variables with missing data: BMI, Smoking status and Alcohol status. Smoking status had the lowest proportion of missing data with 25.1%, BMI had 33.5% missing and alcohol status had the highest proportion with 39.8%.

Table 3-4: Proportion of missingness for confounding variables with missing data.

Characteristic	Missing Observation N (%)
BMI	5,509 (33.5)
Smoking	4,124 (25.1)
Alcohol	6,531 (39.8)

Table 3-5 shows the baseline characteristics of a complete case dataset versus all eligible patients.

Table 3-5: Baseline characteristics of confounding variables with complete case analysis or all eligible patients in the study.

Characteristic		Complete Case	All Eligible Patients
BMI	Observed Mean (SD)	28.8 (±4.8)	28.7 (±4.9)
	Missing N (%)	-	5,509 (33.5)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**
Age	40 to <60	646 (9.7)	2,158 (13.1)
	60 to <70	1,286 (19.3)	2,928 (17.8)
	70 to <80	2,504 (37.6)	5,399 (32.9)
	80 and older	2,220 (33.4)	5,948 (36.2)
Gender	Male	2,626 (39.5)	6,467 (39.4)
	Female	4,030 (60.5)	9,966 (60.6)
Alcohol	No	4,842 (72.8)	7,252 (44.1)
	Yes/Ex	1,814 (27.2)	2,650 (16.1)
	Missing	-	6,531 (39.8)
Smoking	Never	5,002 (75.2)	9,188 (55.9)
	Current/Ex	1,654 (24.8)	3,121 (19.0)
	Missing	-	4,124 (25.1)
Charlson Score	0	2,610 (39.2)	7,337 (44.6)
	1	233 (3.5)	767 (5.7)
	≥2	3,813 (57.3)	8,329 (50.7)
Socio-economic status	1	1,064 (16.0)	2,774 (16.9)
	2	1,194 (17.9)	2,773 (16.8)
	3	1,172 (17.6)	2,768 (16.8)
	4	1,227 (18.4)	2,770 (16.9)
	5	1,259 (18.9)	2,770 (16.9)
	Rural	740 (11.2)	2,578 (15.7)
CVD	No	6,061 (91.1)	15,207 (92.5)
	Yes	595 (8.9)	1,226 (7.5)
AF	No	6,087 (91.5)	15,238 (92.7)
	Yes	569 (8.5)	1,195 (7.3)
Hypertension	No	2,099 (31.5)	7,391 (45.0)
	Yes	4,557 (68.5)	9,042 (55.0)
Diabetes	No	5,003 (75.2)	13,442 (81.8)
	Yes	1,653 (24.8)	2,991 (18.2)
Depression/Anxiety	No	5,190 (78.0)	13,223 (80.5)
	Yes	1,466 (22.0)	3,210 (19.5)
Cancer	No	6,381 (95.9)	15,737 (95.8)
	Yes	275 (4.1)	696 (4.2)
Indication	Fracture	2,003 (30.1)	5,200 (31.6)
	Osteoarthritis	4,152 (62.4)	10,017 (61.0)
	Other	501 (7.5)	1,216 (7.4)
Previous Fixation	No	6,260 (94.1)	15,644 (95.2)
	Yes	396 (5.9)	789 (4.8)
Previous Fracture	No	5,512 (82.8)	13,945 (84.9)
	Yes	1,144 (17.2)	2,488 (15.1)

Auxiliary Variables

There were insufficient observations to run all variables within the cohort. Variables included in exploratory models were reduced for between-term collinearity or for likelihood that it is captured elsewhere. For example, the use of hypertensive medication is captured by patients diagnosed with hypertensive disease. A similar scenario for Rheumatoid Arthritis and Diabetes existed. The diagnosis variable in these cases is used as there are patients with hypertension who do not use medication thus more patients are captured. The variables investigated included:

Co-morbidities measured prior to surgery

- Rheumatoid Arthritis
- Other inflammatory disease
- Other inflammatory arthritis
- Osteoarthritis
- Paget's Disease
- Dementia
- Chronic Kidney Disease
- Chronic Obstructive Pulmonary Disease (COPD)
- Asthma
- Peptic ulcer
- Dyslipidaemia

Pharmaceutical use

- Antibiotics
- Anticonvulsants
- Anticoagulation prior to surgery
- Bisphosphonates
- Hypnotic medication (Benzodiazepine)
- Bone protection (Bisphosphonate, Calcium, Vitamin D etc)
- Corticosteroids

For smoking status and alcohol status only one variable each was shown to possibly be associated with smoking in addition to those in the main analysis: the presence of COPD prior to surgery for smoking and the presence of dyslipidaemia for alcohol.

For BMI, no further auxiliary variables were identified beyond those already included in the main analysis. Following stepwise regression modelling only the COPD present prior to surgery variable was included as an auxiliary variable within the multiple imputation model (Table 3-6). Other clinically potential auxiliary variables were excluded due to between-term collinearity when running the stepwise model or found to be above the removal threshold.

Table 3-6: Auxiliary variables showing statistically significant association with confounding variable containing missing data.

Variable	Auxiliary Variable	p-value
Smoking	Sex	0.000
	COPD	0.000
	Age	0.000
Alcohol	Sex	0.000
	Physio Referral	0.008
	Indication for surgery	0.018
	Dyslipidaemia	0.009
BMI	Age	0.000
	Hypertension	0.005
	Indication for surgery	0.002

Plausibility of MAR assumption

Table 3-7, Table 3-8 and Table 3-9 show the characteristics of the main analysis model of patients with missing data against those with complete data. Several variables were identified as potentially explanatory for use of imputing missing values under the MAR assumption and therefore are not MCAR. For BMI, potential variables with higher proportions (approximately 10%) of BMI observed values included: age, Charlson score, cardiovascular disease, hypertension, diabetes, NSAID use, and GP visits. For alcohol status potential variables included Charlson score, atrial fibrillation, hypertension, diabetes, depression, Orthopaedic referrals, physiotherapy referrals and revision. For smoking status potential variables included Charlson score, hypertension, diabetes, NSAID use, GP visits, Orthopaedic referrals, and physiotherapy referrals.

Table 3-7: Proportion of patients with a BMI, alcohol, or smoking values and those with missing values for each potential confounder explanatory variable.

Confounders	BMI Value %	BMI Missing %	Alcohol Value %	Alcohol Missing %	Smoking Value %	Smoking Missing %
Age						
40 to <60	55.7	44.3	50.1	49.1	71.7	28.3
60 to <70	70.5	29.5	63.1	36.9	77.4	22.6
70 to <80	73.0	27.0	64.1	35.9	77.7	22.3
80 and older	62.5	37.5	59.0	41.0	72.3	27.7
Sex						
Male	66.5	33.5	59.8	40.2	75.8	24.2
Female	66.5	33.5	60.6	39.4	74.3	25.7
Charlson score						
0	61.5	38.5	55.8	44.2	70.8	29.2
1	48.1	51.9	47.3	56.7	66.4	33.6
≥2	72.6	27.4	65.4	34.6	79.3	20.7
Socio-economic status						
1	62.4	37.6	61.5	38.5	72.7	27.3
2	67.0	33.0	62.4	37.6	77.8	22.2
3	67.3	32.7	61.5	38.5	75.7	24.5
4	70.3	29.7	63.5	36.5	77.0	23.0
5	72.1	27.9	62.6	37.4	80.5	19.5
6	59.2	40.8	49.3	50.7	65.2	34.8

Cardiovascular disease						
No	65.8	34.2	59.5	40.5	74.4	25.6
Yes	74.6	25.4	69.2	30.8	80.7	19.3
Atrial Fibrillation						
No	65.9	34.1	59.5	40.5	74.5	25.5
Yes	73.3	26.7	70.0	30.0	79.8	20.2
Hypertension						
No	51.5	48.5	50.5	49.5	66.3	33.7
Yes	78.8	21.2	68.3	31.7	82.0	18.0
Diabetes						
No	62.5	37.5	58.0	42.0	73.0	27.0
Yes	84.3	15.7	70.5	29.5	83.6	16.4
Depression/ Anxiety						
No	65.0	35.0	58.4	41.6	74.0	26.0
Yes	72.7	27.3	67.9	32.1	78.9	21.1
Cancer						
No	66.5	33.5	60.4	39.6	74.9	25.1
Yes	65.7	34.3	56.0	44.0	74.7	25.3
Indication for surgery						
Fracture	60.0	40.0	54.8	45.2	77.0	23.0
Osteoarthritis	70.2	29.8	63.0	37.0	73.3	26.7
Other	63.9	36.1	60.8	39.2	78.8	21.2
Previous fixation						
No	66.2	33.8	60.0	40.0	74.5	25.5
Yes	73.0	27.0	66.3	33.7	82.6	17.4
Previous fracture						
No	65.8	34.2	59.3	40.7	73.7	26.3
Yes	70.3	33.5	65.5	34.5	81.6	18.4

Table 3-8: Proportion of patients with a BMI, alcohol, or smoking values and those with missing values for each potential exposure explanatory variable.

Exposures	BMI Value %	BMI Missing %	Smoking Value %	Smoking Missing %	Alcohol Value %	Alcohol Missing %
NSAIDs use						
Quintile 1	56.3	43.7	57.1	42.9	53.1	46.9
Quintile 2	64.4	35.6	57.5	42.5	78.8	21.2
Quintile 3	68.8	31.2	59.3	40.7	81.4	18.6
Quintile 4	72.0	28.0	64.3	35.7	84.4	15.6
Quintile 5	73.8	26.2	63.2	36.8	86.0	14.0
Atypical Analgesic use						
No	66.3	33.7	60.6	39.4	74.6	25.4
Yes	69.2	30.8	53.7	46.3	80.2	19.8
Tramadol use						
No	65.3	34.7	60.2	39.8	73.2	26.8
Yes	71.4	28.6	60.7	39.3	81.7	18.3
GP visits						
1-4	61.4	38.6	57.4	42.6	76.5	23.5
0	50.2	49.8	53.8	46.2	31.0	69.0
>4	72.7	27.3	63.1	36.9	85.6	14.4
Orthopaedic referral						
No	65.9	34.1	59.4	40.6	73.9	26.1
Yes	74.0	26.0	71.4	28.6	87.7	12.3
Physiotherapy referral						
No	65.7	34.3	59.1	40.9	73.8	26.2
Yes	74.7	25.3	72.5	27.5	86.2	13.8

Table 3-9: Proportion of patients with a BMI, alcohol, or smoking values and those with missing values for the potential outcome explanatory variable.

Outcome	BMI Value %	BMI Missing %	Smoking Value %	Smoking Missing %	Alcohol Value %	Alcohol Missing %
Revision						
No	68.5	31.5	63.1	36.9	75.5	24.5
Yes	68.4	31.6	42.7	56.3	79.8	20.2

Backward stepwise regression of all variables against missingness of BMI, smoking and alcohol confirmed few variables could explain missingness (Table 3-10).

Table 3-10: Backward stepwise regression against missingness of BMI, alcohol, and smoking variables.

Variable	Explanatory Variable	p-value
BMI	Diabetes	0.038
	Sex	0.036
	NSAIDs	0.053
	Cardiovascular disease	0.006
	Atrial fibrillation	0.001
	Hypertension	0.000
Smoking	Previous fracture	0.029
	Depression/Anxiety	0.024
	Socio-economic status	0.029
Alcohol	Sex	0.040

Characteristics following multiple imputation

Baseline characteristics before multiple imputation was compared to post multiple imputation for variables imputed (Table 3-11). All other variables remained the same and will be tabulated later in this chapter.

Table 3-11: pre-and post-multiple imputation baseline characteristics for BMI, smoking and alcohol variables.

Characteristic	Category	Pre-MI	Post-MI
BMI	Observed Mean (SD)	28.7 (±4.9)	28.2 (±4.9)
		Observation N (%)	Observation N (%)
Alcohol	No	7,252 (73.2)	12,013 (73.1)
	Yes/Ex	2,650 (26.8)	4,420 (26.9)
Smoking	Never	9,188 (76.3)	12,280 (74.7)
	Current/Ex	3,121 (23.7)	4,153 (25.3)

Figure 3-14 shows the distribution of BMI, smoking status and alcohol status as imputed in each of the 40 datasets.

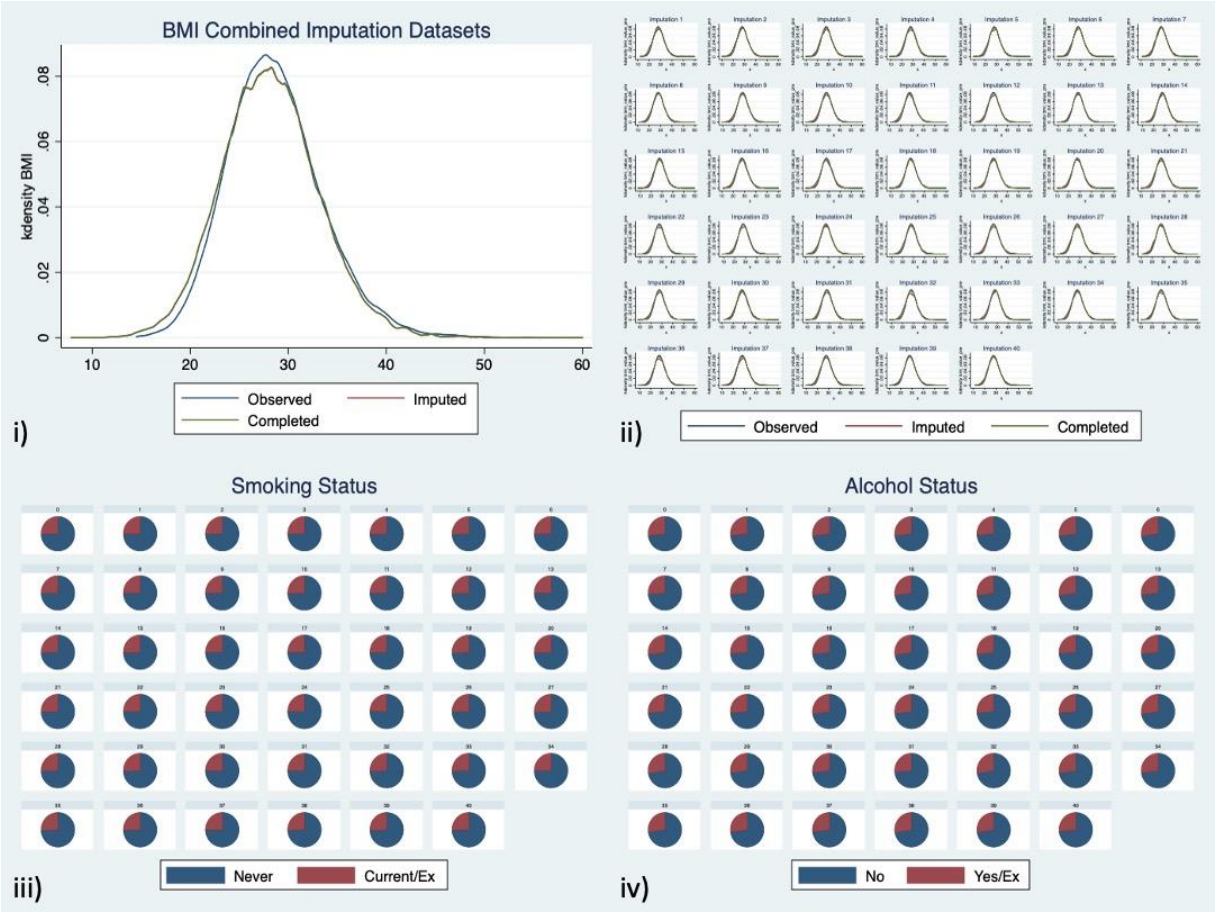


Figure 3-14: Distribution of missing variables following multiple imputation. i) BMI histogram within combined datasets. ii) BMI histogram within each dataset. iii) Smoking distribution within each dataset. iv) Alcohol distribution within each dataset.

Evaluation of multiple imputation

Subsequent experimental chapters will report the pre- and post-multiple imputation univariate regression models and the MC errors.

3.4.3. Study population and baseline characteristics

Data from the RACat registry was found for 17,034 who underwent primary Total Hip Arthroplasty (THA) surgery from 2005 to July 2012. Of these 270 (1.6%) had revision surgery within the first post-operative year and 331 (1.9%) were under the age of 40 years. These patients were excluded leaving our study population as 16,433 (Figure 3-15).

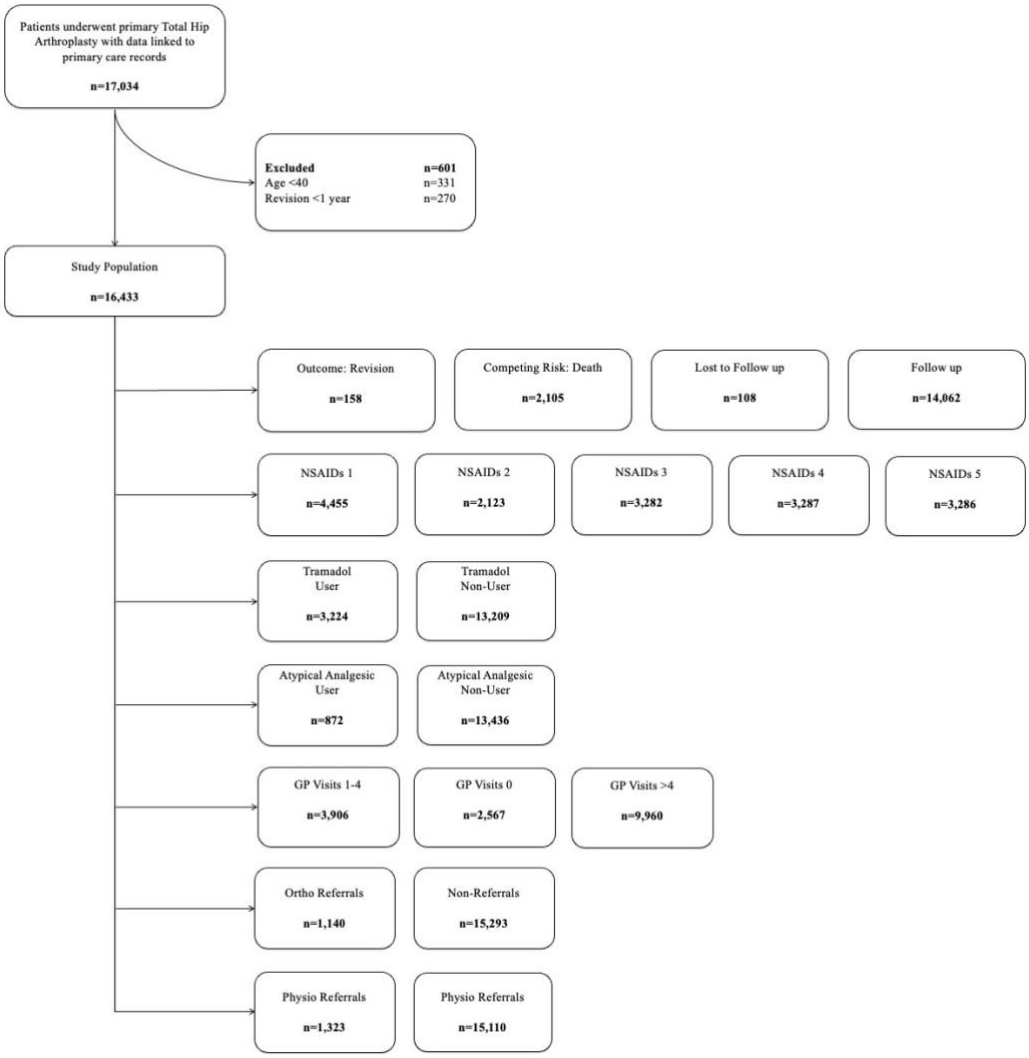


Figure 3-15: Patient selection flow-chart.

Baseline characteristics for the study population following multiple imputation can be found in Table 3-12, and for each exposure in Appendix 5.

Table 3-12: Baseline characteristics following multiple imputation.

Characteristic		
BMI	Observed Mean (SD)	28.2 (±4.9)

Characteristic	Category	Observation N (%)
Age	40 to <60	2,158 (13.1)
	60 to <70	2,928 (17.8)
	70 to <80	5,399 (32.9)
	80 and older	5,948 (36.2)
Gender	Male	6,467 (39.4)
	Female	9,966 (60.6)
Alcohol	No	12,013 (73.1)
	Yes/Ex	4,420 (26.9)
Smoking	Never	12,280 (74.7)
	Current/Ex	4,153 (25.3)
Charlson Score	0	7,337 (44.6)
	1	767 (5.7)
	≥2	8,329 (50.7)
Socio-economic status	1	2,774 (16.9)
	2	2,773 (16.8)
	3	2,768 (16.8)
	4	2,770 (16.9)
	5	2,770 (16.9)
	Rural	2,578 (15.7)
CVD	No	15,207 (92.5)
	Yes	1,226 (7.5)
AF	No	15,238 (92.7)
	Yes	1,195 (7.3)
Hypertension	No	7,391 (45.0)
	Yes	9,042 (55.0)
Diabetes	No	13,442 (81.8)
	Yes	2,991 (18.2)
Depression/Anxiety	No	13,223 (80.5)
	Yes	3,210 (19.5)
Cancer	No	15,737 (95.8)
	Yes	696 (4.2)
Indication	Fracture	5,200 (31.6)
	Osteoarthritis	10,017 (61.0)
	Other	1,216 (7.4)
Previous Fixation	No	15,644 (95.2)
	Yes	789 (4.8)
Previous Fracture	No	13,945 (84.9)
	Yes	2,488 (15.1)

3.4.4. Follow-up

Participants were followed-up for a median (inter-quartile range) of 2.25 (1.22 to 3.74) years. During this time 158 (0.96%) patients underwent revision surgery, 108 (0.66%) were lost to follow-up and 2,105 (12.8%) died.

A lifetable depicting estimated numbers at risk, numbers revised, numbers withdrawn, and survival rates can be found in Table 3-13.

Table 3-13: Lifetable of the outcome of interest (revision).

Year	Numbers at Risk	Numbers Revised	Numbers Withdrawn	Survival %	95% Confidence Intervals
1	16,433	0*	3,296	100	
2	13,137	89	3,933	99.2	99.0-99.4
3	9,115	27	3,109	98.9	98.6-99.0
4	5,979	20	2,226	98.4	98.1-98.7
5	3,733	15	1,628	97.9	97.5-98.3
6	2,090	5	1,214	97.6	97.1-98.0
7	871	2	676	97.2	96.5-97.9
8	193	0	193	97.2	96.5-97.9
***Excluded**					

3.4.5. Rates of revision

NSAID use

NSAID use was not uncommon in the first post-operative year with Daily Defined Dose percentiles at 20, 40, 60 and 80 of 0, 13.3, 46.7 and 104.0 respectively.

Increasing NSAID use within the first-year post operation was associated with increasing incidence of revision, ranging from 3.07/1000 (2.06-4.58) among the lowest quintile of NSAID use to 5.78/1000 (4.46-7.49) in the top quintile (Table 3-14). The median (inter-quartile range) of time to revision per quintile is shown in Table 3-14. This finding is confirmed in the Kaplan-Meier estimates of cumulative risk of revision. (Figure 3-16).

Table 3-14: Unadjusted rates of revision for the NSAID exposure.

NSAID Quintile	Number (%)	Number of Revisions	Median (inter-quartile range) of years to revision	Revision rate /1,000 person-years (95% CI)
1	4,455 (27.1)	24	1.9 (1.5-3.8)	3.07 (2.06-4.58)
2	2,123 (12.9)	14	2.3 (1.8-2.8)	2.36 (1.39-3.98)
3	3,282 (20.0)	27	2.2 (1.3-4.1)	2.87 (1.97-4.19)
4	3,287 (20.0)	36	1.6 (1.3-2.9)	3.71 (2.68-5.15)
5	3,286 (20.0)	57	1.9 (1.6-2.8)	5.78 (4.46-7.49)

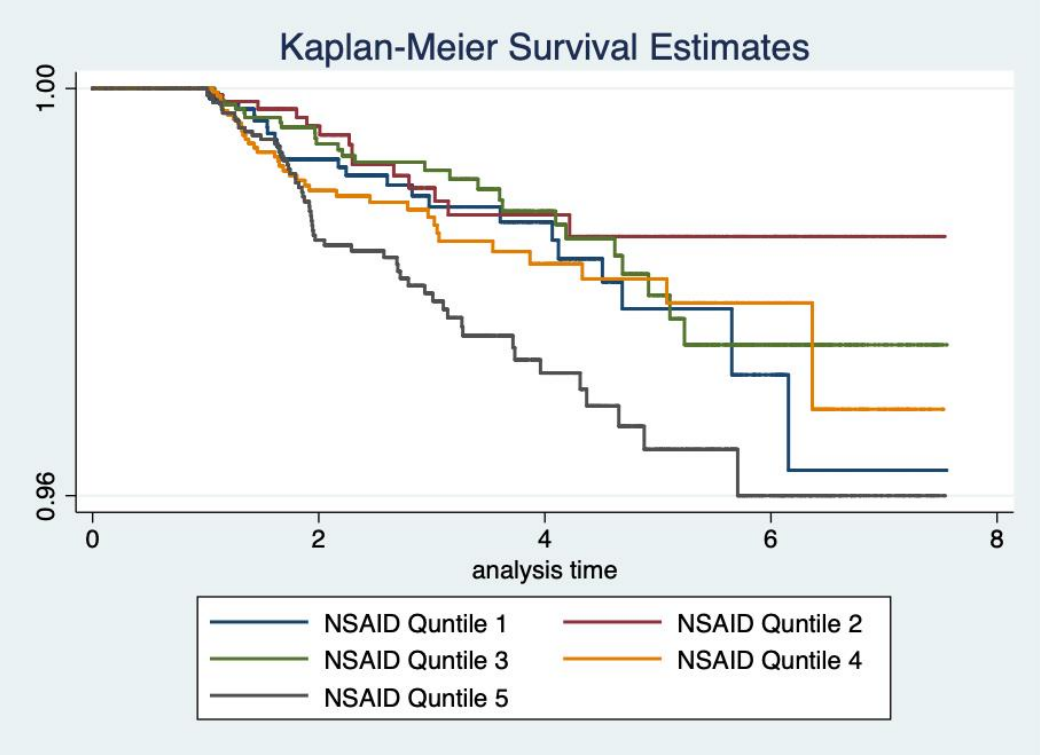


Figure 3-16: Kaplan-Meier survival estimates for the NSAID exposure.

Tramadol use

Tramadol use was also not uncommon in the first post-operative year with 3,224 (19.6%) patients identified as users.

Tramadol use within the first-year post operation was associated with an increased incidence of revision, 6.43/1000 (5.08-8.14) compared to non-user with 2.78/1000 (2.56-3.42) (Table 3-15). The median (inter-quartile range) of time to revision is shown in Table 3-15. This finding is confirmed in the Kaplan-Meier estimates of cumulative risk of revision (Figure 3-17).

Table 3-15: Unadjusted rates of revision for the Tramadol exposure.

Tramadol Users	Number (%)	Number of Revisions	Median (inter-quartile range) of years to revision	Revision rate /1,000 person-years (95% CI)
No	13,209 (80.4)	89	1.9 (1.3-3.0)	2.78 (2.56-3.42)
Yes	3,224 (19.6)	69	1.9 (1.6-3.1)	6.43 (5.08-8.14)

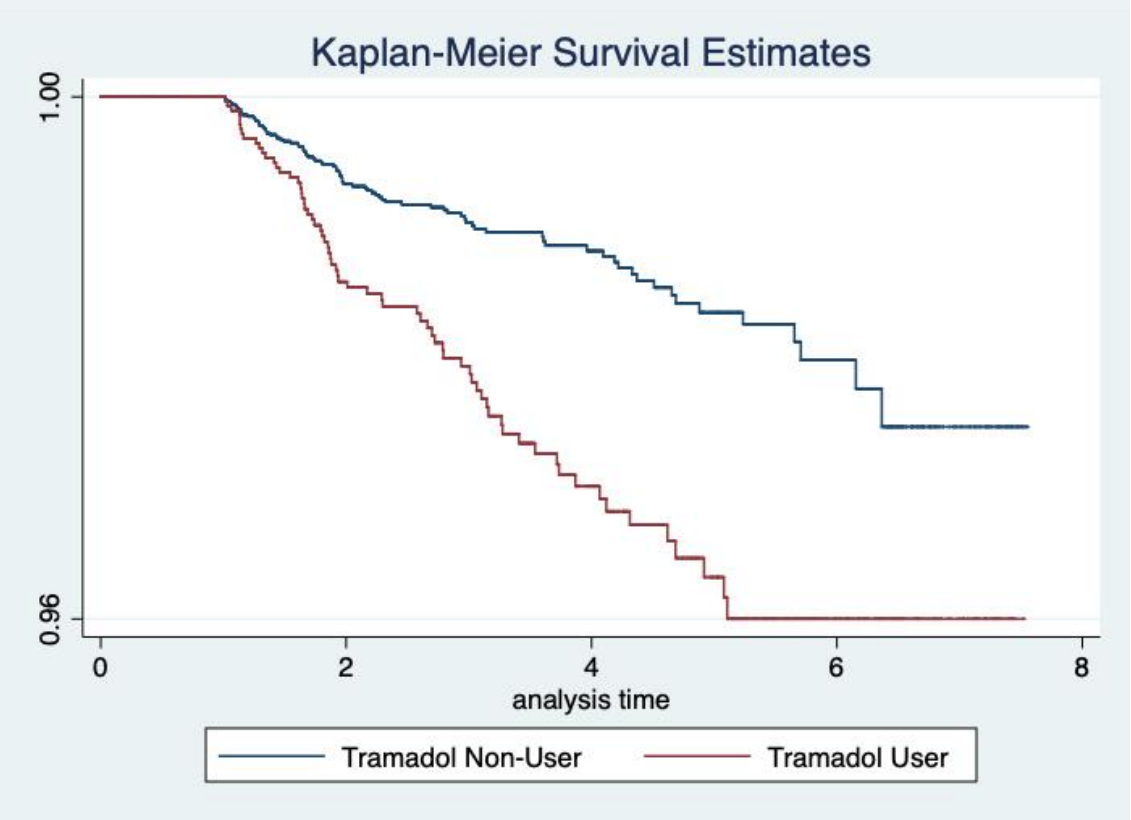


Figure 3-17: Kaplan-Meier survival estimates for the Tramadol exposure.

Atypical analgesic use

Atypical analgesic use was not common for new users in the first post-operative year with 872 (6.1%) patients identified.

Atypical analgesic use within the first-year post operation was associated with an increased incidence of revision, 8.70/1000 (6.04-12.51) compared to non-user with 3.25/1000 (2.70-3.91) (Table 3-16). The median (inter-quartile range) of time to revision is shown in Table 3-16. This finding is confirmed in the Kaplan-Meier estimates of cumulative risk of revision. (Figure 3-18).

Table 3-16: Unadjusted rates of revision for the Atypical Analgesic exposure.

Atypicals Users	Number (%)	Number of Revisions	Median (inter-quartile range) of years to revision	Revision rate /1,000 person-years (95% CI)
No	13,436 (93.9)	113	1.9 (1.3-2.8)	3.25 (2.70-3.91)
Yes	872 (6.1)	29	2.6 (1.7-3.3)	8.70 (6.04-12.51)

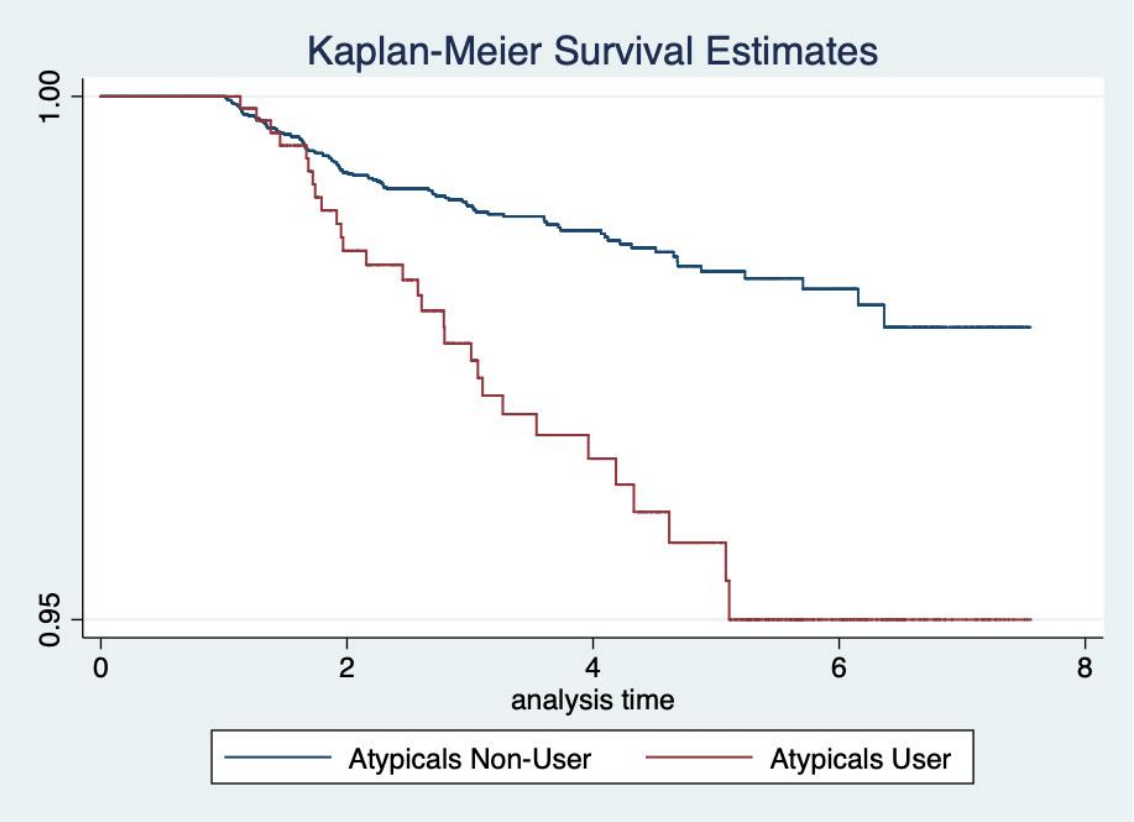


Figure 3-18: Kaplan-Meier survival estimates for the Atypical Analgesic exposure.

GP visits

A large proportion of patients attended the GP surgery more than 4 times per year in the first post-operative year with 9,960 (60.6%) being identified.

Attending more than four times within the first-year post operation was associated with an increased incidence of revision, 4.17/1000 (3.50-4.97) compared to patients seen at least once and less than or equal to 4 times with 2.02/1000 (1.29-3.16) (Table 3-17). The median (inter-quartile range) of time to revision is shown in Table 3-17. This finding is confirmed in the Kaplan-Meier estimates of cumulative risk of revision. (Figure 3-19).

Table 3-17: Unadjusted rates of revision for the frequency of GP visits exposure.

GP Visits	Number (%)	Number of Revisions	Median (inter-quartile range) of years to revision	Revision rate /1,000 person-years (95% CI)
1-4	3,906 (23.8)	19	2.0 (1.3-2.9)	2.02 (1.29-3.16)
0	2,567 (15.6)	14	2.1 (1.5-2.8)	4.21 (2.50-7.11)
>4	9,960 (60.6)	125	1.9 (1.4-3.0)	4.17 (3.50-4.97)

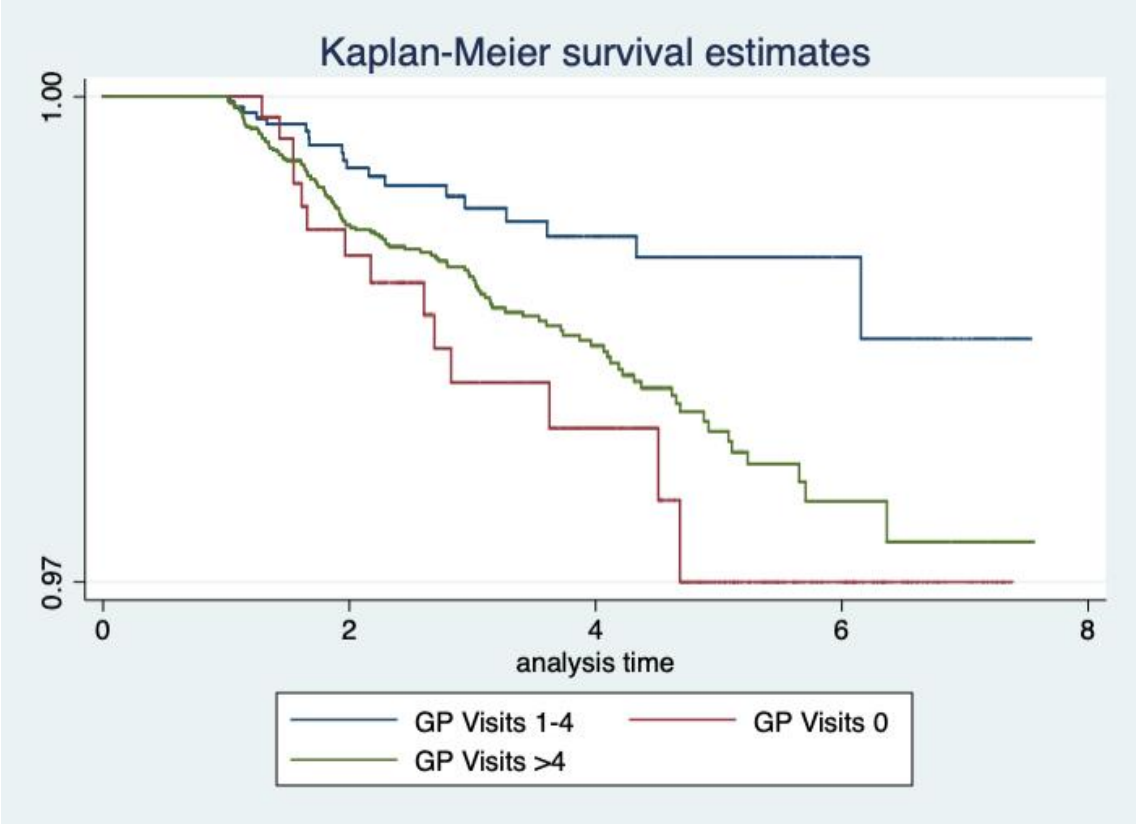


Figure 3-19: Kaplan-Meier survival estimates for the frequency of GP visits exposure.

Orthopaedic referral

The number of patients referred to the Orthopaedic department within the first post-operative year was 1,140 (6.9%).

Being referred to the Orthopaedic department within the first-year post operation was associated with an increased incidence of revision, 6.51/1000 (4.10-10.34) compared to patients not referred with 3.50/1000 (2.97-4.13) (Table 3-18). The median (inter-quartile range) of time to revision is shown in Table 3-18. This finding is confirmed in the Kaplan-Meier estimates of cumulative risk of revision. (Figure 3-20).

Table 3-18: Unadjusted rates of revision for the referral to the Orthopaedic department exposure.

Ortho Referral	Number (%)	Number of Revisions	Median (inter-quartile range) of years to revision	Revision rate /1,000 person-years (95% CI)
No	15,293 (93.1)	140	1.9 (1.0-6.4)	3.50 (2.97-4.13)
Yes	1,140 (6.9)	18	1.9 (1.1-3.2)	6.51 (4.10-10.34)

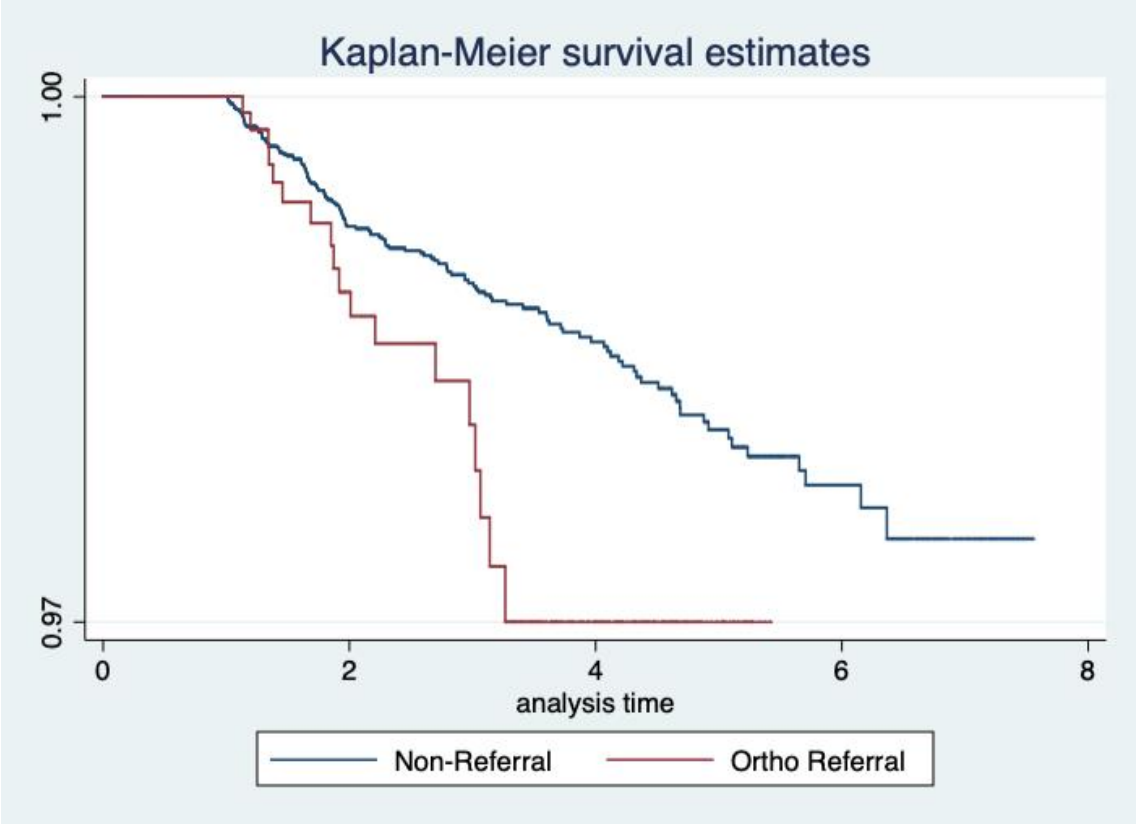


Figure 3-20: Kaplan-Meier survival estimates for the referral to the Orthopaedic department exposure.

Physiotherapy referral

The number of patients referred to the Physiotherapy department within the first post-operative year was 1,323 (8.1%).

Being referred to the Physiotherapy department within the first-year post operation was associated with an increased incidence of revision, 6.51/1000 (4.24-9.98) compared to patients not referred with 3.47/1000 (2.93-4.10) (Table 3-19). The median (inter-quartile range) of time to revision is shown in Table 3-19. This finding is confirmed in the Kaplan-Meier estimates of cumulative risk of revision. (Figure 3-21).

Table 3-19: Unadjusted rates of revision for the referral to the Physiotherapy department exposure.

Physio Referral	Number (%)	Number of Revisions	Median (inter-quartile range) of years to revision	Revision rate /1,000 person-years (95% CI)
No	15,110 (91.9)	137	1.9 (1.0-6.4)	3.44 (2.92-4.07)
Yes	1,323 (8.1)	21	1.7 (1.0-4.7)	7.61 (4.61-10.83)

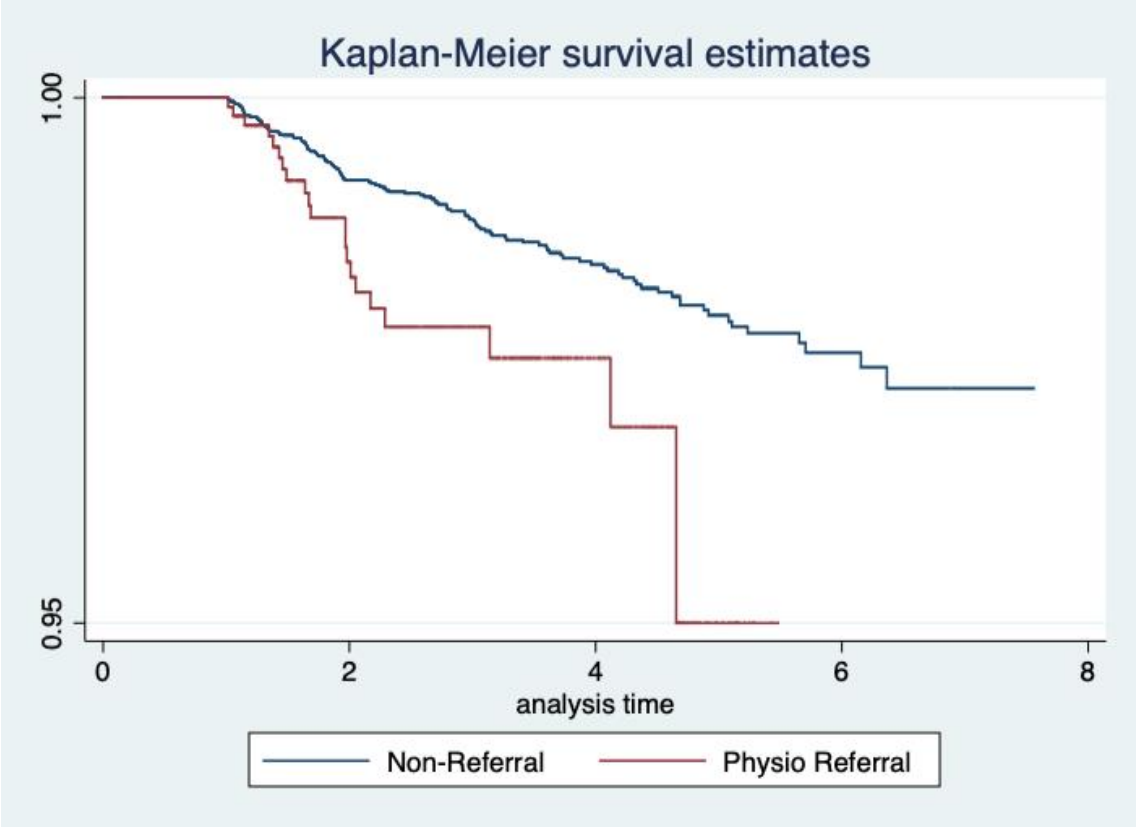


Figure 3-21: Kaplan-Meier survival estimates for the referral to the Physiotherapy department exposure.

3.4.6. Reasons for revision surgery

Reasons recorded for revision surgery can be found in Table 3-20.

Table 3-20: Reasons for revision surgery.

Reason for revision	Observation N (%)
Hip fracture	4 (2.5)
Osteoarthritis	1 (0.6)
Congenital bone disease	1 (0.6)
Acquired bone disease	4 (2.5)
Mechanical complication	87 (55.1)
Prosthesis infection	27 (17.1)
Missing	34 (21.6)

3.4.7. Exposure versus mortality

There was a statistically significant association between each exposure and mortality within the cohort, suggesting that mortality would need to be accounted for as a competing risk in the regression using Fine & Gray modelling (Table 3-21).

Table 3-21: Univariate and multivariate Cox proportional hazards regression for the analgesic and healthcare utilisation exposures against mortality.

	Univariate HR (95% CI)	Multivariable HR (95% CI)
Second Quintile compared to First Quintile (p<20)	0.57 (0.05-0.64) **p<0.001**	0.65 (0.57-0.75) **p<0.001**
Third Quintile compared to First Quintile (p<20)	0.46 (0.41-0.52) **p<0.001**	0.58 (0.52-0.66) **p<0.001**
Fourth Quintile compared to First Quintile (p<20)	0.33 (0.29-0.38) **p<0.001**	0.42 (0.37-0.48) **p<0.001**
Fifth Quintile (p≥80) compared to First Quintile (p<20)	0.30 (0.27-0.35) **p<0.001**	0.42 (0.37-0.49) **p<0.001**
Tramadol Users vs. Non-Users	0.56 (0.49-0.63) **p<0.001**	0.70 (0.62-0.80) **p<0.001**
Atypical Users vs. Non-Users	0.53 (0.42-0.65) **p<0.001**	0.69 (0.55-0.86) **p<0.001**
GP Visits 0 vs. GP Visits 1-4	1.82 (1.61-2.06) **p<0.001**	1.42 (1.25-1.61) **p<0.001**
GP Visits >4 vs. GP Visits 1-4	0.50 (0.46-0.55) **p<0.001**	0.51 (0.46-0.57) **p<0.001**
Non-Referral vs. Ortho Referral	0.40 (0.31-0.51) **p<0.001**	0.45 (0.35-0.59) **p<0.001**
Non-Referral vs. Physio Referral	0.52 (0.42-0.64) **p<0.001**	0.50 (0.40-0.61) **p<0.001**

3.4.8. Main outcome

Survival analysis demonstrated rates of revision after the first-year post operation were positively associated with all exposures.

THA participants in the top quintile (p≥80) compared to the lowest quintile (p<20) corresponded to an adjusted Sub-Hazard Ratio (SHR) of 2.20 [1.35-3.59; p=0.002]. THA participants who used Tramadol compared to non-users corresponded to an adjusted SHR of 2.42 [1.74-3.35; p<0.001]. THA participants who used atypical analgesics compared to non-users corresponded to an adjusted SHR of 2.59 [1.70-3.95; p<0.001]. THA participants who attended the surgery more than four times in the first post-operative year compared to those who attended 1-4 times corresponded to an adjusted SHR of 2.10 [1.29-3.42; p=0.003]. THA participants who were referred to the Orthopaedic department in the first post-operative year compared to those who were not referred corresponded to an adjusted SHR of 1.90 [1.16-3.13; p=0.011]. THA participants who were referred to the Physiotherapy department in the first post-operative year compared to those who were not referred corresponded to an adjusted SHR of 2.07 [1.29-3.32; p=0.003] (Table 3-22).

Table 3-22: Univariate and multivariate Fine and Gray (F&G) analysis to investigate the association with analgesic and healthcare utilisation within the first post-operative year and subsequent revision rate.

	Unadjusted SHR (95% CI)	Multivariable SHR (95% CI)
Second Quintile compared to First Quintile (p<20)	0.80 (0.41-1.55) p=0.511	0.78 (0.40-1.51) p=0.458
Third Quintile compared to First Quintile (p<20)	1.00 (0.58-1.73) p=0.997	1.00 (0.58-1.73) p=0.995
Fourth Quintile compared to First Quintile (p<20)	1.32 (0.79-2.21) p=0.297	1.41 (0.83-2.39) p=0.206
Fifth Quintile (p≥80) compared to First Quintile (p<20)	2.03 (1.26-3.27) **p=0.003**	2.20 (1.35-3.59) **p=0.002**
Tramadol Users vs. Non-Users	2.36 (1.72-3.23) **p<0.001**	2.42 (1.74-3.35) **p<0.001**
Atypical Users vs. Non-Users	2.76 (1.84-4.13) **p<0.001**	2.59 (1.70-3.95) **p<0.001**
GP Visits 0 vs. GP Visits 1-4	1.98 (1.00-3.95) p=0.052	2.07 (1.02-4.18) **p=0.043**
GP Visits >4 vs. GP Visits 1-4	2.12 (1.31-3.34) p=0.997	2.10 (1.29-3.42) **p=0.003**
Non-Referral vs. Ortho Referral	1.95 (1.19-3.20) **p=0.008**	1.90 (1.16-3.13) **p=0.011**
Non-Referral vs. Physio Referral	2.06 (1.30-3.27) **p=0.002**	2.07 (1.29-3.32) **p=0.003**

Fine & Gray regression compared to Cox proportional regression

For NSAID, Tramadol and Atypical analgesic use a 12.8%, 4.3% and 5.3% difference in the main outcome between the multivariate Cox and Fine & Gray models were found respectively. For GP visits and referrals to the Orthopaedic or Physiotherapy department a 12.1% (visits equal to zero) and 7.8% (visits greater than four), 3.3% and 2.5% difference in the main outcome between the multivariate Cox and Fine & Gray models were found respectively (Table 3-23).

Table 3-23: Comparison of Cox proportional hazard regression and Fine and Gray (F&G) analysis of the association between analgesic and healthcare utilisation and subsequent revision rate.

	Cox Multivariable (95% CI)	F&G Multivariable (95% CI)
Second Quintile compared to First Quintile (p<20)	0.71 (0.37-1.39) p=0.321	0.78 (0.40-1.51) p=0.458
Third Quintile compared to First Quintile (p<20)	0.91 (0.52-1.59) p=0.743	1.00 (0.58-1.73) p=0.995
Fourth Quintile compared to First Quintile (p<20)	1.25 (0.73-2.12) p=0.416	1.41 (0.83-2.39) p=0.206
Fifth Quintile (p≥80) compared to First Quintile (p<20)	1.95 (1.18-3.21) **p=0.009**	2.20 (1.35-3.59) **p=0.002**
Tramadol Users vs. Non-Users	2.32 (1.68-3.21) **p<0.001**	2.42 (1.74-3.35) **p<0.001**
Atypical Users vs. Non-Users	2.46 (1.63-3.73) **p<0.001**	2.59 (1.70-3.95) **p<0.001**
GP Visits 0 vs. GP Visits 1-4	2.32 (1.15-4.69) **p=0.019**	2.07 (1.02-4.18) **p=0.043**
GP Visits >4 vs. GP Visits 1-4	1.95 (1.20-3.18) **p=0.007**	2.10 (1.29-3.42) **p=0.003**
Non-Referral vs. Ortho Referral	1.84 (1.12-3.02) **p=0.016**	1.90 (1.16-3.13) **p=0.011**
Non-Referral vs. Physio Referral	2.02 (1.26-3.22) **p=0.003**	2.07 (1.29-3.32) **p=0.003**

3.4.9. Post-estimation analysis

NSAID use

Roctab function in Stata was used to calculate general sensitivity and specificity. Sensitivity was 58.86% and Specificity was 60.18%. A Receiver Operating Curve (ROC) revealed an area under the curve of 0.62 (Figure 3-22).

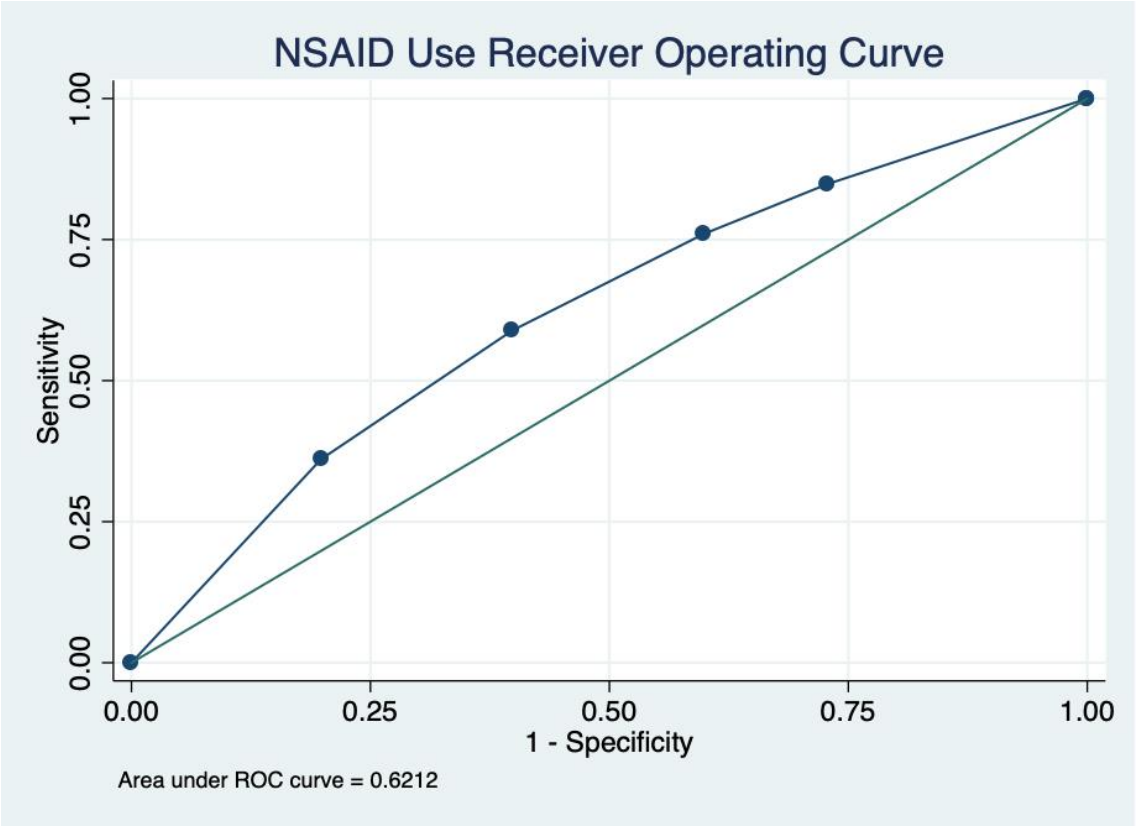


Figure 3-22: Receiver operator curve (ROC) and area under the ROC curve for the NSAID exposure.

The Positive Predictive Value (PPV) was calculated for the top quintile as 1.7% (57/3,286). Negative Predictive Value (NPV) was 99.2% (13,046/13,147). Sensitivity was 36.1% (57/158) and Specificity was 80.2% (13,046/16,275). Table 3-24.

Table 3-24: Sensitivity and specificity table with comparisons between the top quintile of NSAID use and other quintiles combined within the first post-operative year.

	Revision Positive (%)	**Revision Negative (%)**	**Total (%)**
NSAID 5th Quintile	57 (36.1)	3,229 (19.8)	**3,286**
NSAID 1st – 4th Quintiles	101 (63.9)	13,046 (80.2)	**13,147**
Total	**158**	**16,275**	

Using the "tvc" option in Stata, the proportional sub-hazard assumption was tested showing no significant difference in the sub-hazard ratio over time with a p-value of 0.268.

This finding is confirmed with the Cumulative Incidence Function graph comparing the top quintile with the lowest (Figure 3-23).

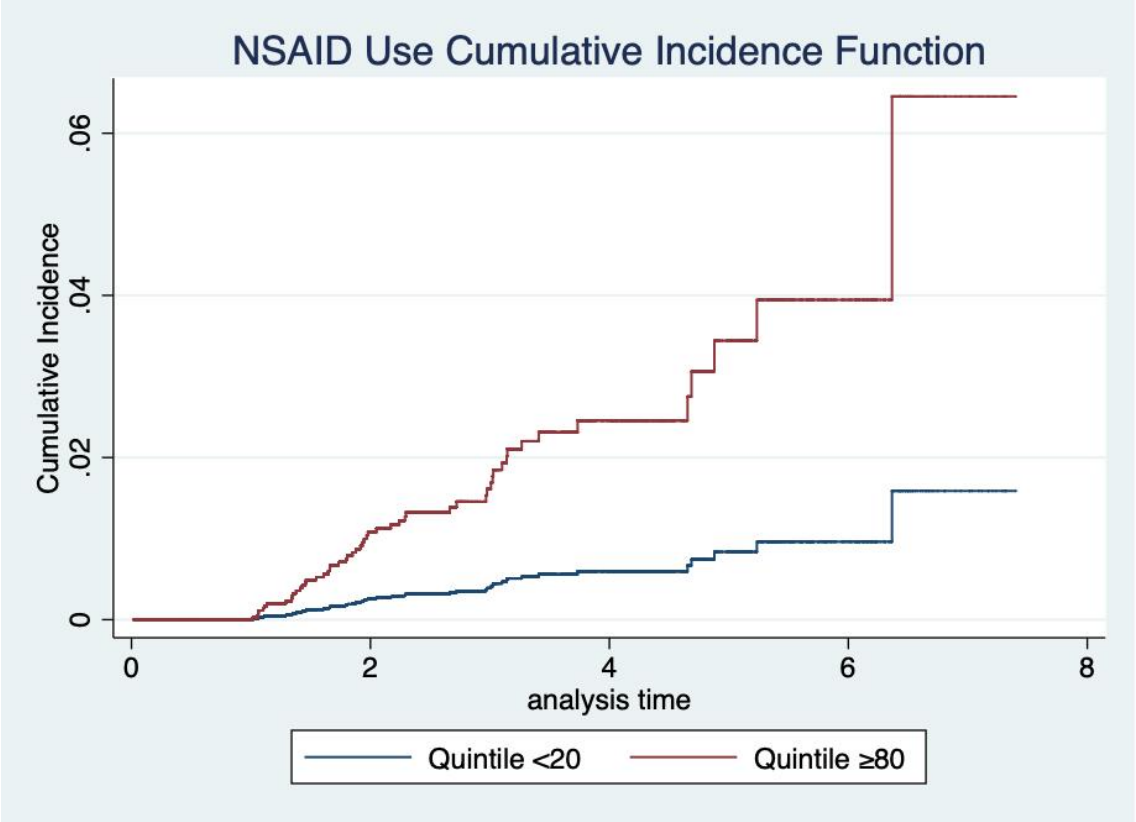


Figure 3-23: Cumulative incidence function for the NSAID exposure comparing the top quintile and other quintiles combined.

Tramadol use

Sensitivity was 43.67% and Specificity was 80.61%. A Receiver Operating Curve (ROC) revealed an area under the curve of 0.62 (Figure 3-24).

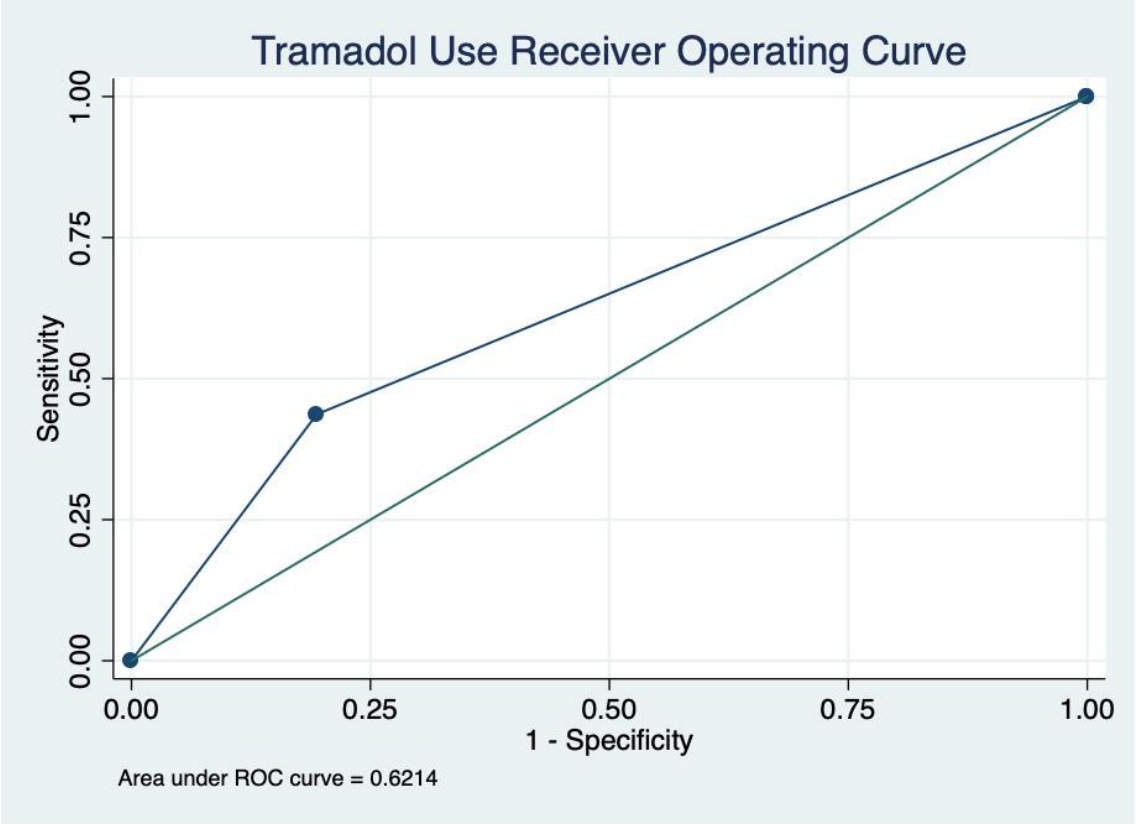


Figure 3-24: Receiver operator curve (ROC) and area under the ROC curve for the Tramadol exposure.

The PPV was calculated as 2.1% (69/3,224). NPV was 99.3% (13,120/13,209). Sensitivity was 43.7% (69/158) and Specificity was 80.6% (13,120/16,275). Table 3-25.

Table 3-25: Sensitivity and specificity table with comparisons between Tramadol users and non-users within the first post-operative year.

	Revision Positive (%)	**Revision Negative (%)**	**Total (%)**
Tramadol Users	69 (43.7)	3,155 (19.4)	**3,224**
Tramadol Non-Users	89 (56.3)	13,120 (80.6)	**13,209**
Total	**158**	**16,275**	

The proportional sub-hazard assumption was tested showing no significant difference in the sub-hazard ratio over time with a p-value of 0.642.

This finding is confirmed with the Cumulative Incidence Function graph comparing the top quintile with the lowest (Figure 3-25).

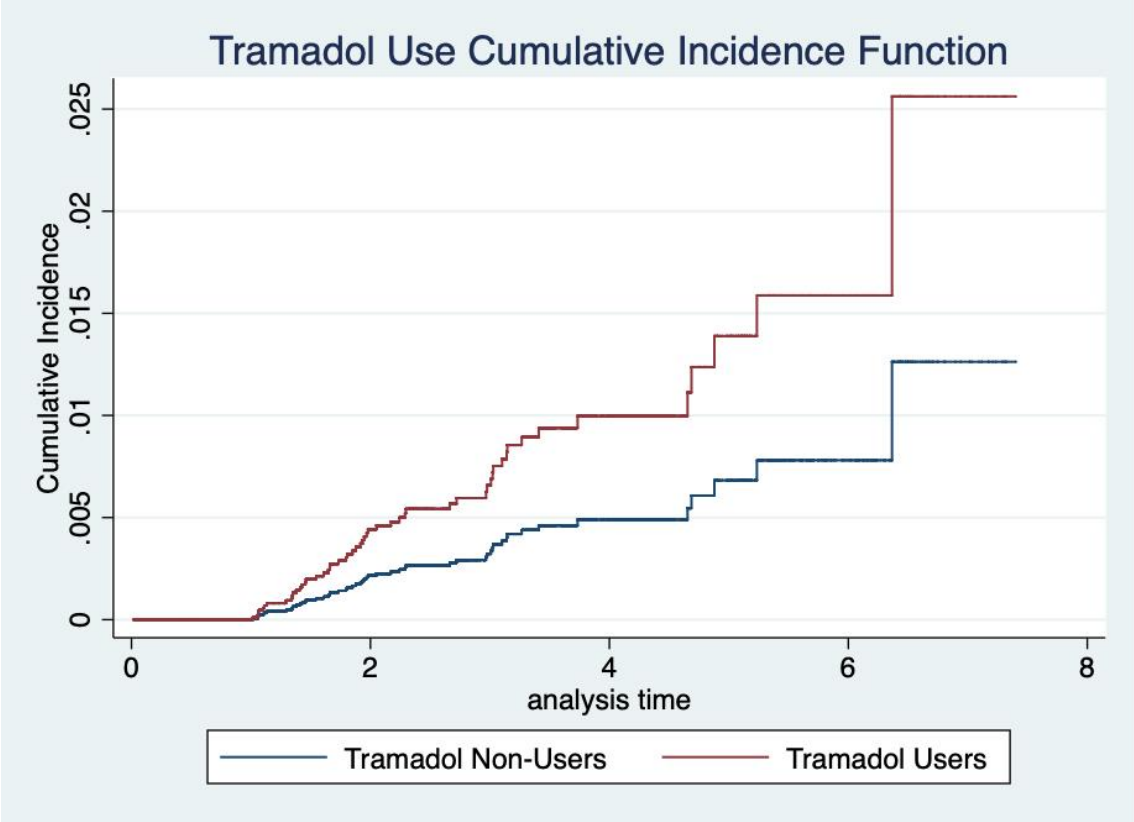


Figure 3-25: Cumulative incidence function for the Tramadol exposure.

Atypical analgesia

Sensitivity was 20.42% and Specificity was 94.05%. A ROC revealed an area under the curve of 0.57 (Figure 3-26).

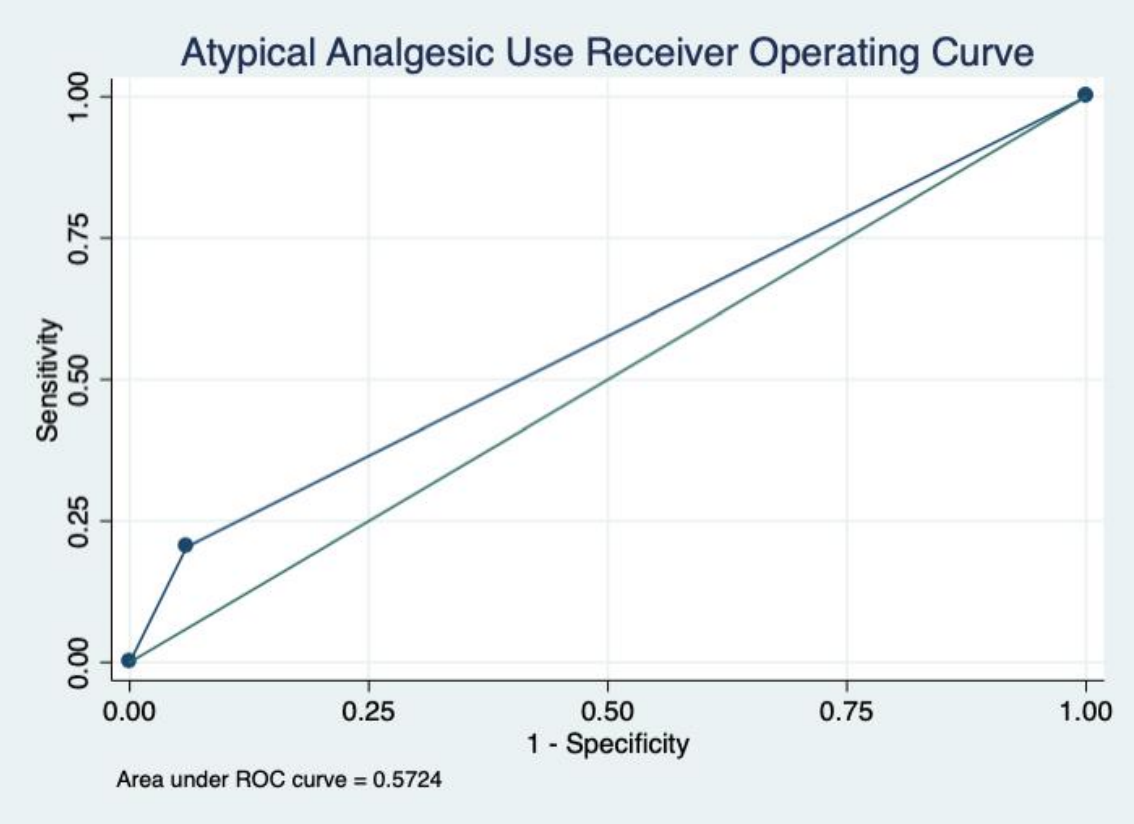


Figure 3-26: Receiver operator curve (ROC) and area under the ROC curve for the Atypical Analgesic exposure.

The PPV was calculated as 3.3% (29/872). NPV was 99.2% (13,323/13,436). Sensitivity was 20.4% (29/142) and Specificity was 94.0% (13,323/14,166). Table 3-26.

Table 3-26: Sensitivity and specificity table with comparisons between Atypical Analgesic users and non-users within the first post-operative year.

	Revision Positive (%)	**Revision Negative (%)**	**Total (%)**
Atypical Users	29 (20.4)	843 (6.0)	**872**
Non-Users	113 (79.6)	13,323 (94.0)	**13,436**
Total	**142**	**14,166**	

The proportional sub-hazard assumption was tested showing no significant difference in the sub-hazard ratio over time with a p-value of 0.306.

This finding is confirmed with the Cumulative Incidence Function graph comparing Atypical analgesic users with non-users (Figure 3-27).

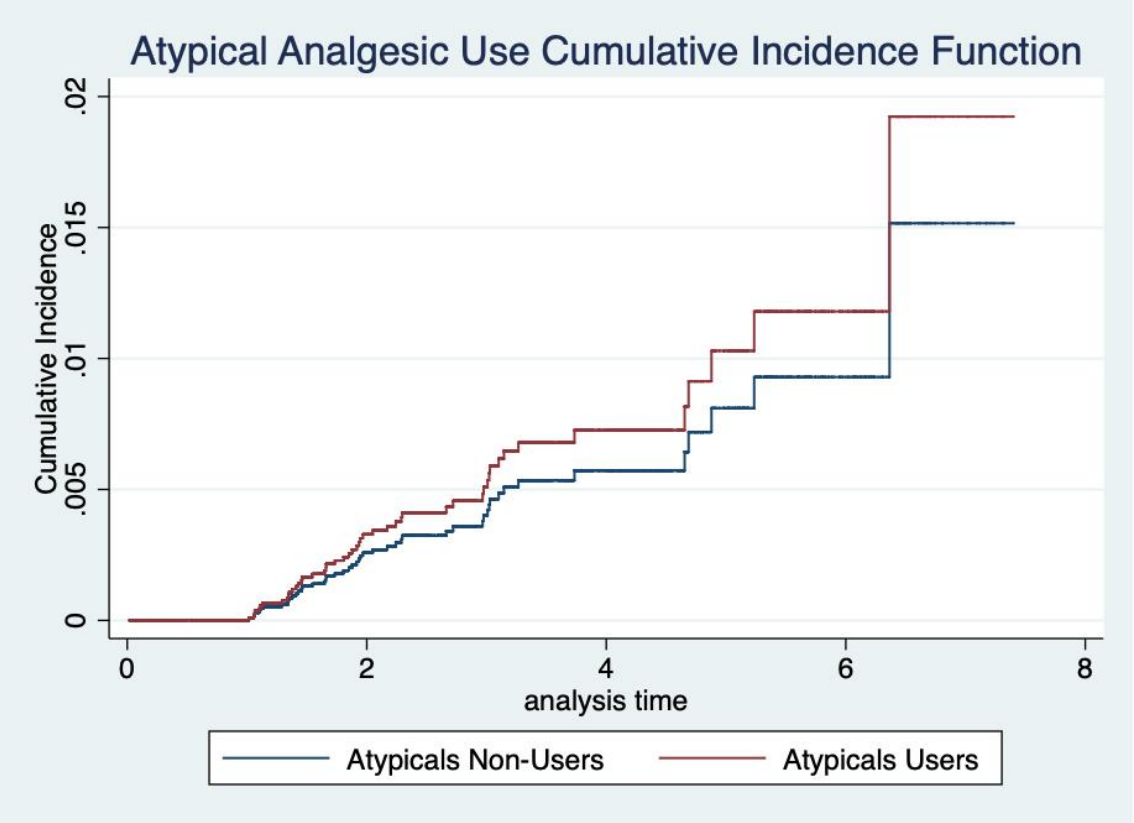


Figure 3-27: Cumulative incidence function for the Atypical Analgesic exposure.

GP visits

Sensitivity was 79.11% and Specificity was 39.57%. A ROC revealed an area under the curve of 0.59 (Figure 3-28).

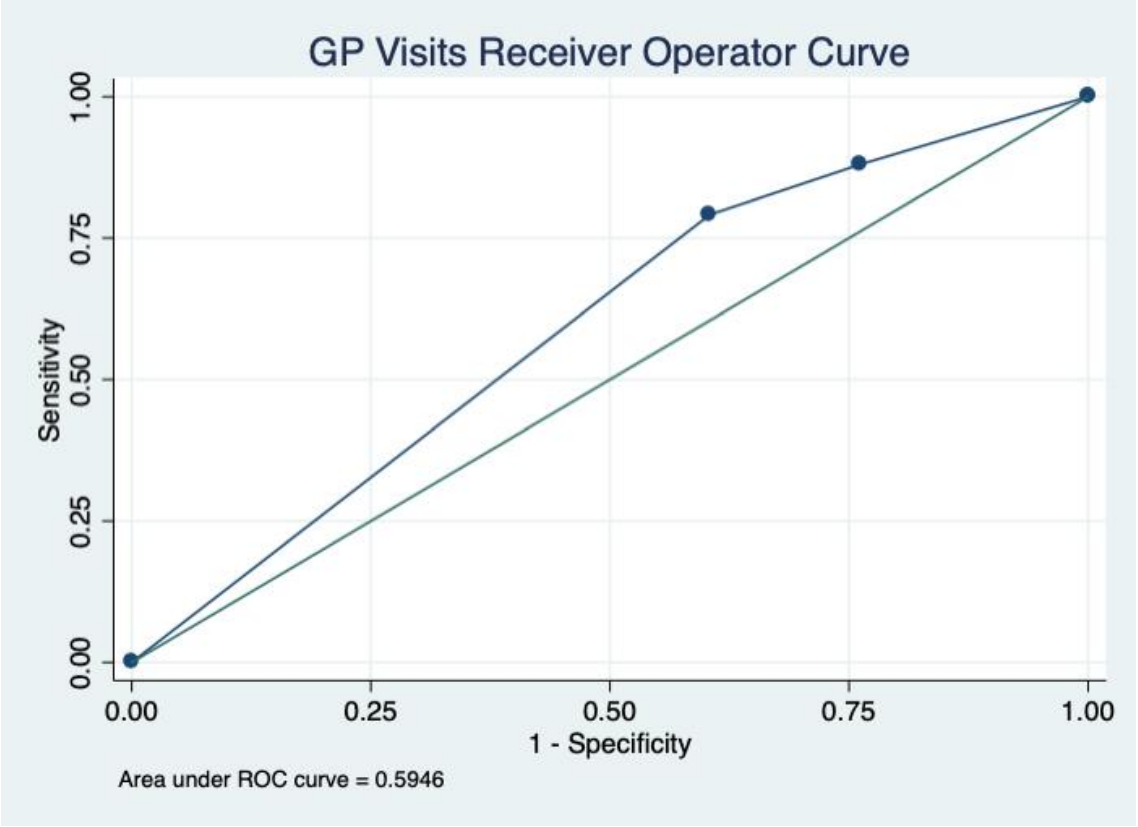


Figure 3-28: Receiver operator curve (ROC) and area under the ROC curve for the frequency of GP visits exposure.

The PPV was calculated for the category GP Visits >4 as 1.3% (125/9,960). NPV was 99.5% (6,440/6,473). Sensitivity was 79.1% (125/158) and Specificity was 39.6% (6,440/16,275). Table 3-27.

Table 3-27: Sensitivity and specificity table with comparisons between patients with more than four GP visits and those with zero to four within the first post-operative year.

	Revision Positive (%)	**Revision Negative (%)**	**Total (%)**
GP Visits >4	125 (79.1)	9,835 (60.4)	**9,960**
GP Visits 0-4	33 (20.3)	6,440 (39.6)	**6,473**
Total	**158**	**16,275**	

The proportional sub-hazard assumption was tested showing no significant difference in the sub-hazard ratio over time with a p-value of 0.765.

This finding is confirmed with the Cumulative Incidence Function graph comparing patients who visited the GP surgery more than four times in the first post-operative year and those who visited less than or equal to four (Figure 3-29).

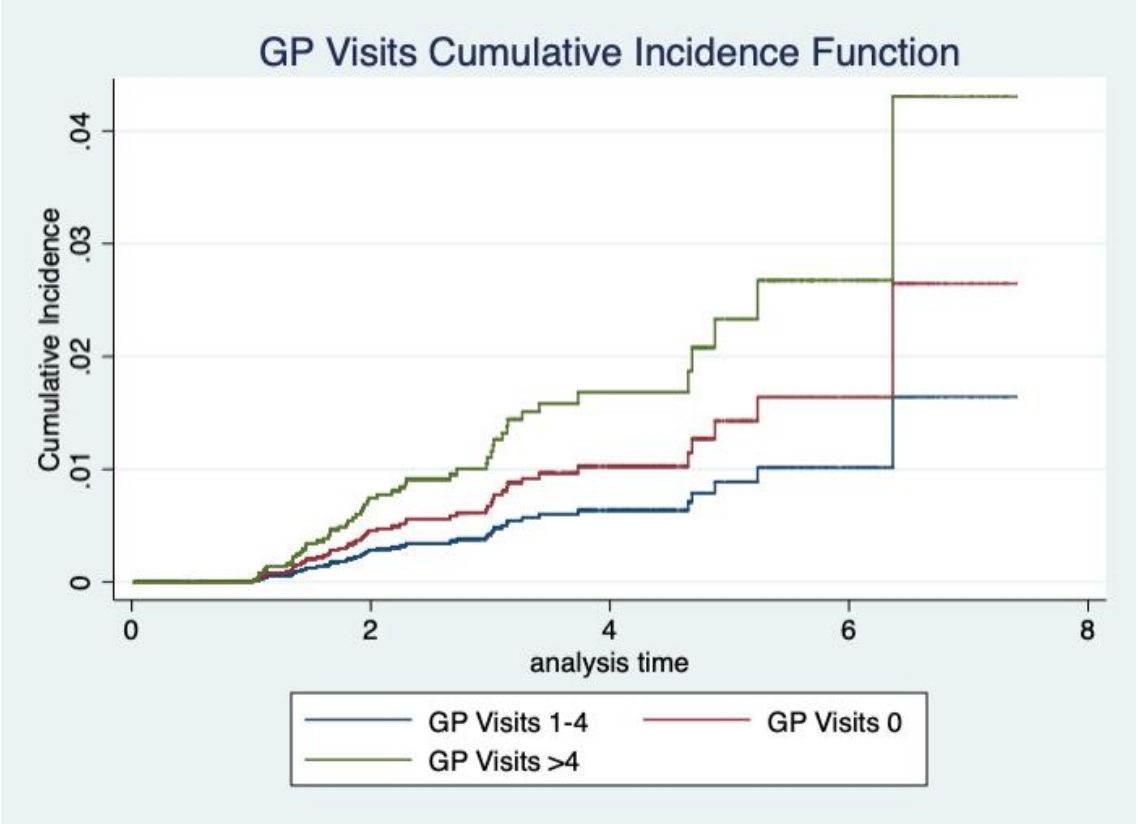


Figure 3-29: Cumulative incidence function for the frequency of GP visits exposure.

Orthopaedic referral

Sensitivity was 11.39% and Specificity was 92.83%. A ROC revealed an area under the curve of 0.52 (Figure 3-30).

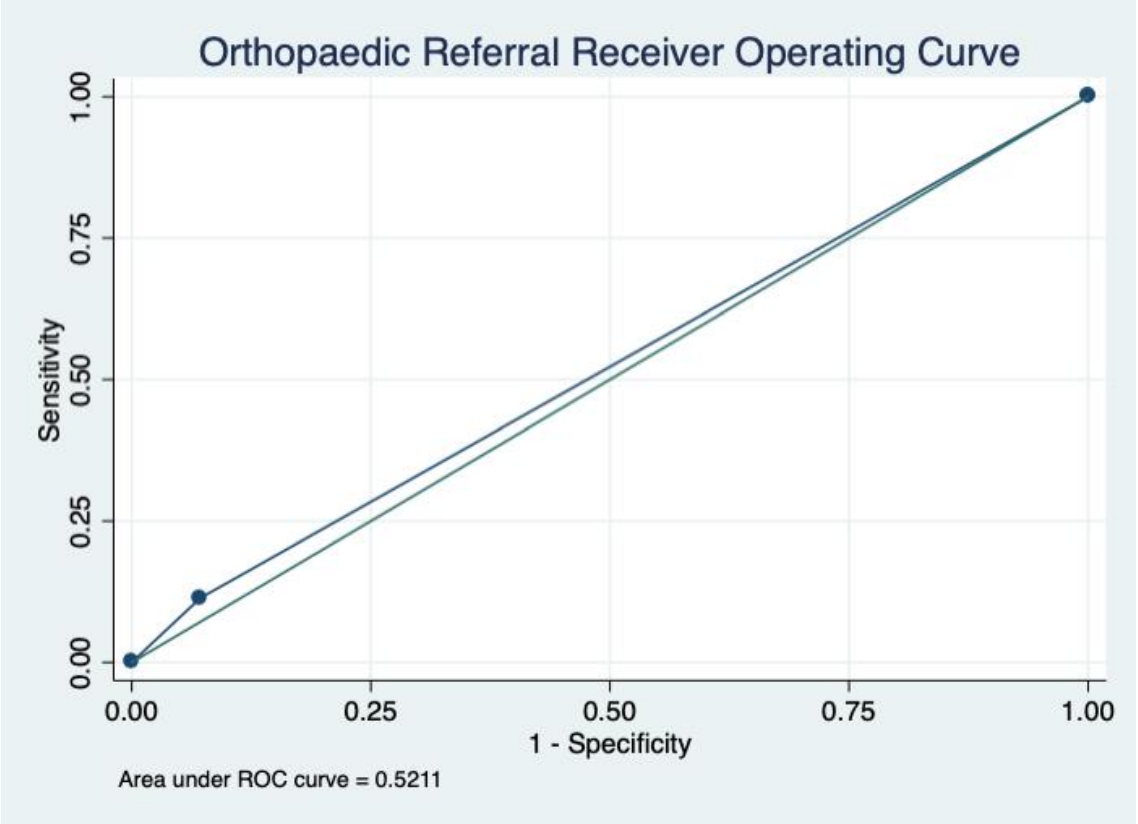


Figure 3-30: Receiver operator curve (ROC) and area under the ROC curve for the referral to the Orthopaedic department exposure.

The PPV was calculated for referred patients as 1.6% (18/1,140). NPV was 99.1% (15,153/15,293). Sensitivity was 11.4% (18/158) and Specificity was 93.1% (15,153/16,275). Table 3-28.

Table 3-28: Sensitivity and specificity table with comparisons between patients referred to the Orthopaedic department and those who were not referred within the first post-operative year.

	Revision Positive	**Revision Negative**	**Total**
Ortho Referral	18	1,122	**1,140**
Non-Referral	140	15,153	**15,293**
Total	**158**	**16,275**	

The proportional sub-hazard assumption was tested showing no significant difference in the sub-hazard ratio over time with a p-value of 0.600.

This finding is confirmed with the Cumulative Incidence Function graph comparing patients referred to the Orthopaedic department in the first post-operative year and those who were not (Figure 3-31).

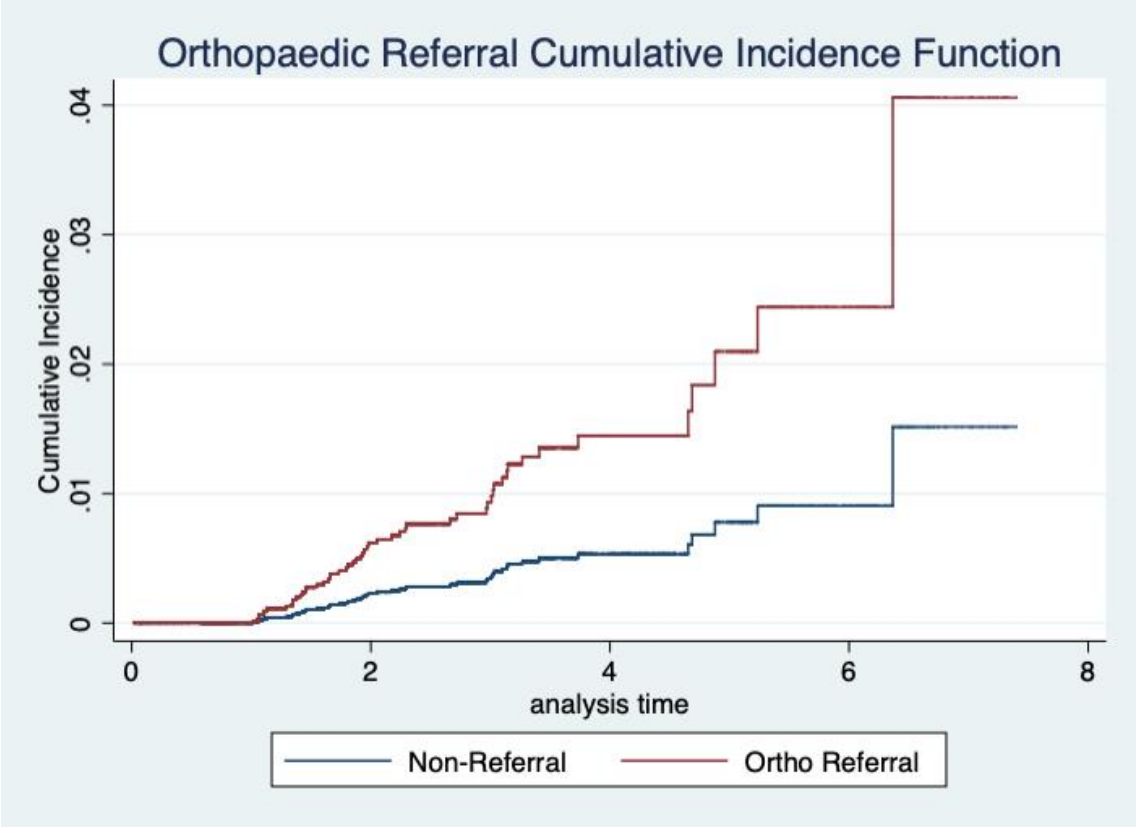


Figure 3-31: Cumulative incidence function for the referral to the Orthopaedic department exposure.

Physiotherapy referral

Sensitivity was 13.29% and Specificity was 91.20%. A ROC revealed an area under the curve of 0.52 (Figure 3-32).

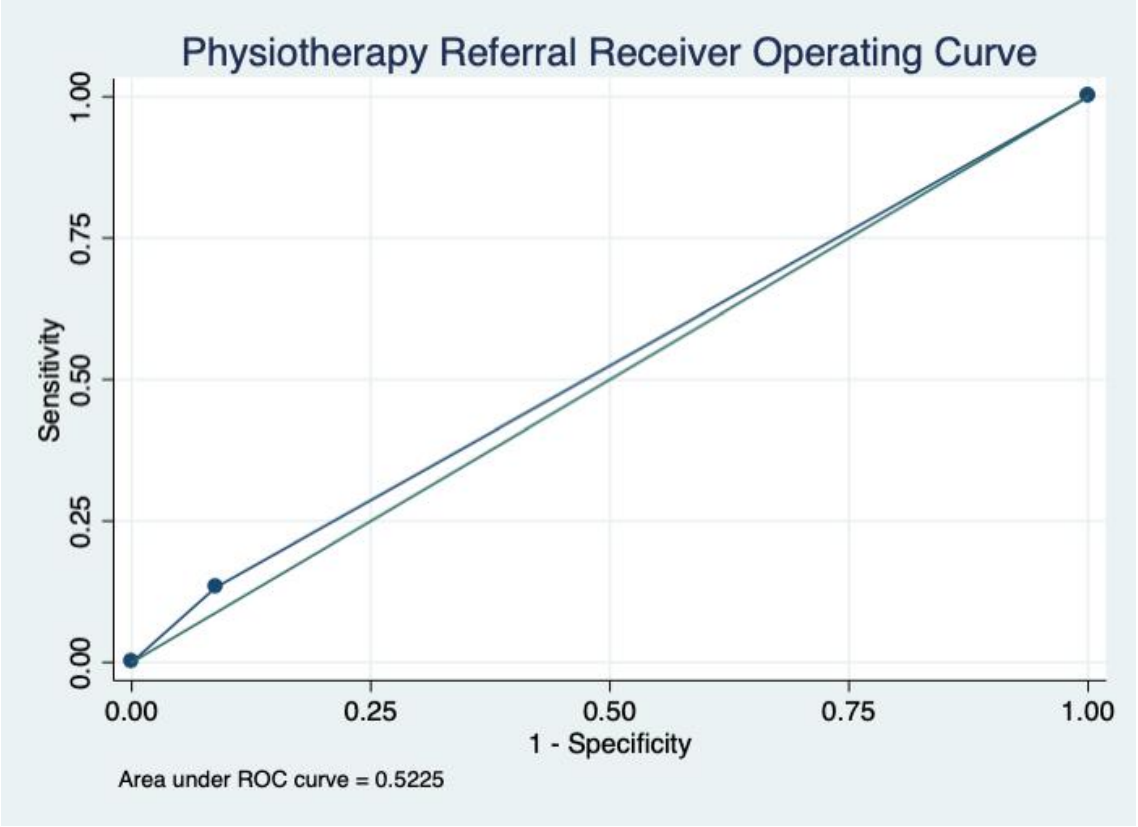


Figure 3-32: Receiver operator curve (ROC) and area under the ROC curve for the referral to the Physiotherapy department exposure.

The PPV was calculated for referred patients as 1.6% (21/1,323). NPV was 99.1% (14,973/15,110). Sensitivity was 13.3% (21/158) and Specificity was 92.0% (14,973/16,275). Table 3-29.

Table 3-29: Sensitivity and specificity table with comparisons between patients referred to the Physiotherapy department and those who were not referred within the first post-operative year.

	Revision Positive	Revision Negative	Total
Physio Referral	21	1,302	**1,323**
Non-Referral	137	14,973	**15,110**
Total	**158**	**16,275**	

The proportional sub-hazard assumption was tested showing no significant difference in the sub-hazard ratio over time with a p-value of 0.982.

This finding is confirmed with the Cumulative Incidence Function graph comparing patients referred to the Physiotherapy department in the first post-operative year and those who were not (Figure 3-33).

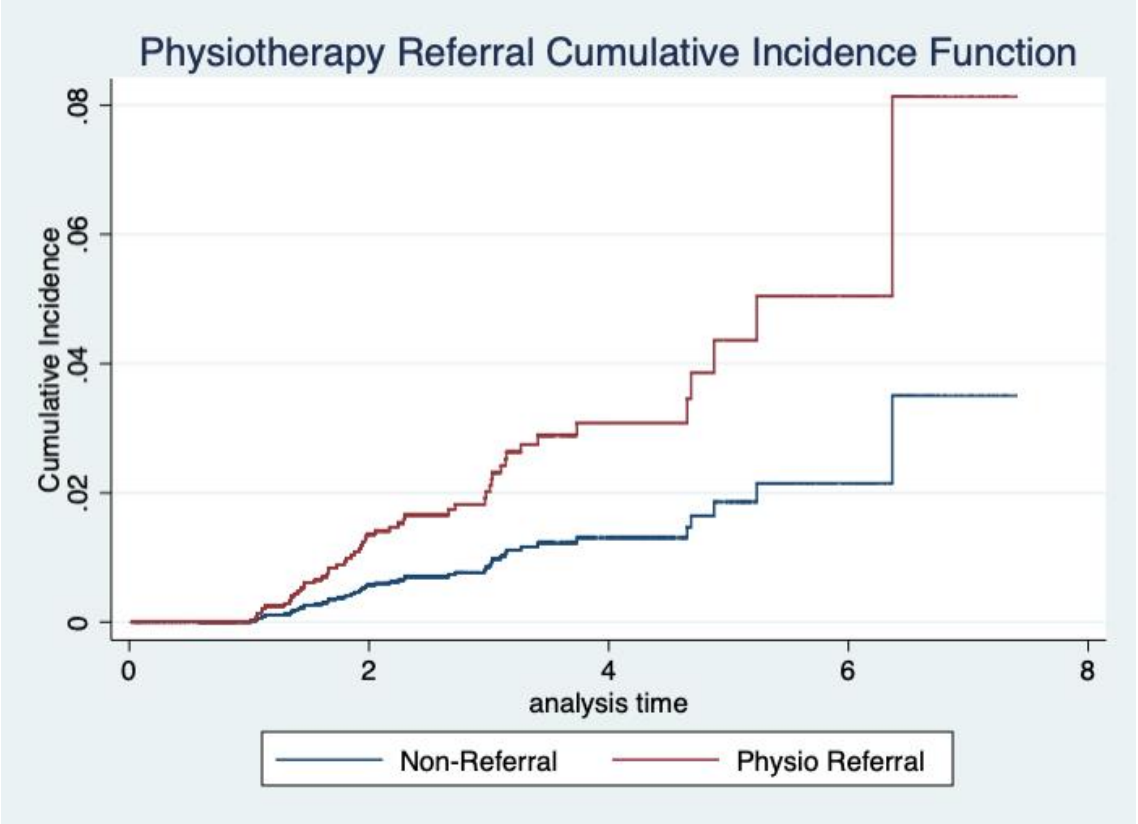


Figure 3-33: Cumulative incidence function for the referral to the Physiotherapy department exposure.

3.4.10. Evaluation of multiple imputation

Identical univariate Cox and F&G modelling was found comparing pre- and post-multiple imputation regressions for all exposures. Comparison tables and Monte Carlo errors can be found in Appendix 6.

3.4.11. Sensitivity analyses

Osteoarthritis as indication only

Following exclusion criteria, 10,017 patients were identified who underwent primary THA for the indication Osteoarthritis (OA) only. Table 3-30 shows revision against OA, bearing type, and fixation type variables.

Table 3-30: Proportion of patients undergoing revision surgery categorised by sensitivity analysis variable.

Characteristic	Category	Revision No	Revision Yes
Osteoarthritis Only	-	9,918	99
Bearing Type	Metal-on-Metal	524	10
	Metal-on-Poly	7,089	71
	Ceramic-on-Ceramic	602	4
	Ceramic-on-Poly	1,125	11
	Missing	6,997	-
Fixation Type	Cemented	3,503	19
	Cementless	7,750	71
	Hybrid	2,353	30
	Missing	2,707	-

Survival analysis demonstrated rates of revision after the first-year post operation were positively associated with exposures in patients with indication for surgery being OA only. THA participants in the top quintile (p≥80) compared to the lowest quintile (p<20) corresponded to an adjusted SHR of 2.73 [1.35-5.51; p=0.005]. THA participants who used Tramadol compared to non-users corresponded to an adjusted SHR of 2.84 [1.88-4.28; p<0.001]. THA participants who used Tramadol compared to non-users corresponded to an adjusted SHR of 2.61 [1.52-4.49; p=0.001]. THA participants who attended the GP surgery more than four times in the first post-operative year

compared to those who attended 1-4 times corresponded to an adjusted SHR of 2.75 [1.37-5.54; p=0.005]. THA participants who were referred to the Orthopaedic department in the first post-operative year compared to those who were not corresponded to an adjusted SHR of 2.25 [1.23-4.14; p=0.009]. THA participants who were referred to the Physiotherapy department in the first post-operative year compared to those who were not corresponded to an adjusted SHR of 2.05 [1.12-3.74; p=0.019].

Differences between fixation types

There were no significant differences between the different fixation types for NSAID use.

Differences between bearing types

There were no significant differences between the different bearing types.

Charlson score exclusion

Associations between exposures and revision rate were not significantly changed when Charlson score was excluded from the analysis (Table 3-31).

Table 3-31: Multivariate Fine and Gray (F&G) analysis to investigate the association with analgesic and healthcare utilisation within the first post-operative year and subsequent revision rate comparing the inclusion or exclusion of Charlson score.

	Multivariable SHR (95% CI) including Charlson score	**Multivariable SHR (95% CI) excluding Charlson score**
Second Quintile compared to First Quintile (p<20)	0.78 (0.40-1.51) p=0.458	0.78 (0.40-1.51) p=0.455
Third Quintile compared to First Quintile (p<20)	1.00 (0.58-1.73) p=0.995	1.00 (0.58-1.73) p=0.998
Fourth Quintile compared to First Quintile (p<20)	1.41 (0.83-2.39) p=0.206	1.41 (0.83-2.40) p=0.205
Fifth Quintile (p≥80) compared to First Quintile (p<20)	2.20 (1.35-3.59) **p=0.002**	2.21 (1.35-3.61) **p=0.002**
Tramadol Users vs. Non-Users	2.42 (1.74-3.35) **p<0.001**	2.43 (1.75-3.37) **p=0.000**
Atypical Users vs. Non-Users	2.59 (1.70-3.95) **p<0.001**	2.60 (1.71-3.97) **p=0.000**
GP Visits 0 vs. GP Visits 1-4	2.07 (1.02-4.18) **p=0.043**	2.09 (1.04-4.21) **p=0.038**
GP Visits >4 vs. GP Visits 1-4	2.10 (1.29-3.42) **p=0.003**	2.11 (1.30-3.43) **p=0.003**
Non-Referral vs. Ortho Referral	1.90 (1.16-3.13) **p=0.011**	1.90 (1.15-3.12) **p=0.012**
Non-Referral vs. Physio Referral	2.07 (1.29-3.32) **p=0.003**	2.06 (1.29-3.30) **p=0.003**

3.4.1. Combined analgesic and healthcare utilisation

Analgesic use and healthcare resource use measured within the first post-operative year were found to be significantly associated with revision rate. However, individually these potential surrogate markers performed poorly when calculating the area under the receiver operating curve. This section investigates these surrogate markers as a combined tool in detecting failing devices.

Logistic regression was performed with all potential surrogate markers combined and odds ratios on a multiple imputation dataset were calculated (Table 3-32).

Table 3-32: Odds ratios following logistic regression for potential surrogate markers combined on a multiple imputation dataset.

	Odds Ratio (95% CI)
NSAID use **Fifth Quintile (p≥80)**	1.51 (0.82-2.80) p=0.187
Tramadol use	2.19 (1.54-3.12) **p<0.001**
Atypical Analgesic use	2.68 (1.75-4.11) **p<0.001**
GP Visits **>4**	1.96 (1.19-3.23) **p<0.001**
Orthopaedic referral	1.19 (0.72-1.99) p=0.495
Physiotherapy referral	1.17 (0.72-1.88) p=0.529

A Receiver Operating Curve revealed an area under the curve of 0.70 for the highest combination of sensitivity and specificity (Figure 3-34).

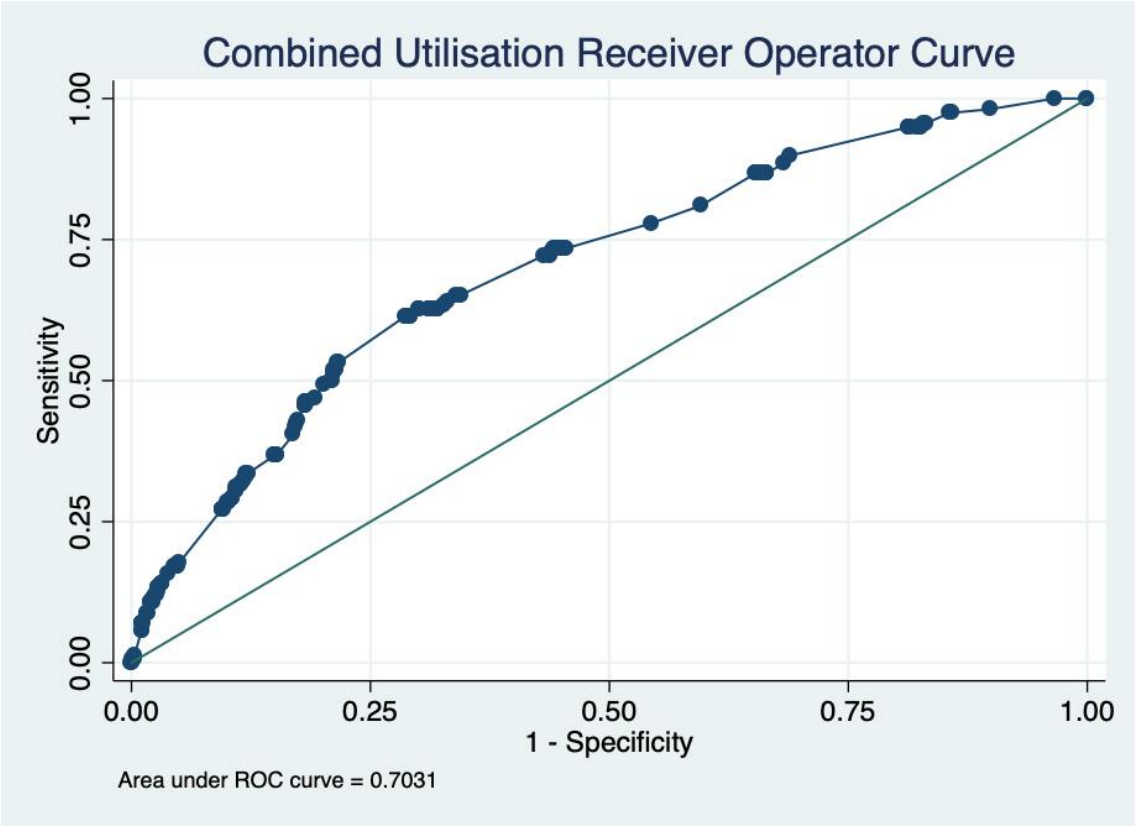


Figure 3-34: Receiver operator curve (ROC) and area under the ROC curve for all exposures combined.

3.5. Discussion

These studies investigated six potential surrogate markers of implant survival in patients undergoing primary THA surgery. This chapter reports analgesic (NSAIDs, Tramadol and Atypical analgesics), and healthcare utilisation (number of visits to primary care and referral to either the Orthopaedic or Physiotherapy department) within the first post-operative year have a significant association with subsequent revision rates. Those patients within the top quintile of NSAID use, who used tramadol or atypical analgesics had more than a two-fold increased risk of revision up to seven years when compared to those requiring doses in the lowest range of NSAID use, non-users of Tramadol or Atypical analgesics. Those patients with more than four visits had approximately a two-fold increased risk of revision up to seven years when compared to those requiring one to four visits. Patients referred to the Orthopaedic department had almost a two-fold increased risk of revision when compared to those not referred. Those patients referred to the Physiotherapy department had over a two-fold increased risk of revision. This suggests that the rate of analgesic or healthcare utilisation may be potential surrogates of hip implant failure allowing for early identification of poorly performing devices. For all but one of the exposures, the risk was further increased following sensitivity analysis investigating patients where the indication for surgery as Osteoarthritis only. Physiotherapy referral maintained a similar risk. The recent failures in Metal-on-Metal implants have highlighted the need for alternative early surrogate markers of outcome[46,59]. The results were therefore stratified to investigate whether analgesic or healthcare utilisation could detect any difference in revision rates between bearings. There was no statistically significant increase in revision rate when results were stratified to Metal-on-Poly (MoP), Ceramic-on-Ceramic (CoC), Ceramic-on-Poly (CoP), and Metal-on-Metal (MoM).

The most common symptom of patients suffering from Osteoarthritis is pain[266], and for those who have undergone Total Hip Arthroplasty (THA) a significant proportion have persistent post-operative pain[251]. Pain is considered an important outcome measure in THA surgery[243] and has been associated with the commonly used endpoint of revision[255]. Monitoring pain has substantial advantages over revision and is a more meaningful measure of success from the patient's perspective. However, its potential use in monitoring implants is highlighted with the fact that it occurs more commonly than revision and thus smaller numbers would be required to detect differences between devices. This is vital when searching for new methods of post-market surveillance where new devices are introduced in small numbers initially with low revision rates[105].

One potential method of surveillance of pain is using Patient Reported Outcome Measures (PROMs) such as the Oxford Hip Score, which have been in use since its introduction in 1996[216]. Its use in post implantation surveillance has increased in England and is now introduced for all hip and knee arthroplasties[566]. It has been demonstrated in the New Zealand Registry that PROMs scores measured at six months is predictive of THA revision rates at two years[174]. Furthermore, 67% of revisions within two years for THR would have been captured by monitoring the lowest 28% of Oxford Hip Scores (OHS)[174]. More recent research has demonstrated an association between PROMS measured at one year post-operatively and late revision rate, and confirmed pain in isolation was most predictive[221].

Analgesic utilisation

Despite the usefulness and validity of PROMs scores as surrogate measures of implant failure, there are multiple limitations including implementation costs, integration with workflow systems, and adequate participation[246,253]. The experiments here support this finding, suggesting that NSAID, Tramadol and atypical analgesic use within the first post-operative year is a potentially valid surrogate of implant revision following primary THA surgery. There are significant advantages to using this method of measuring outcome. The data is collected prospectively and routinely in general practice databases with electronic prescription data and is readily available retrospectively in different countries[567]. This may be an excellent source to monitor signals for early detection of poorly performing devices as defined by outcomes of pain or revision[243]. As pain is commonly treated pharmacologically, monitoring analgesic use would provide considerable data compared to revision rate which is a rare outcome particularly in the early stages following market introduction[255].

In this cohort, 43.7% of patients who subsequently had revision surgery were opioid users in the first post-operative year. Furthermore, 36.1% of patients needing revision surgery were those in the top quintile of NSAID use in the year following surgery. Other studies report 42% of patients requiring revision were opioid users at one year, and over half of those revised reported pain post-operatively[243,255]. This study was unable to assess whether the amount of opioid use was associated with revision rate. However, other studies have demonstrated the quantity of opioid use was higher in patients requiring early revision but not significantly later[243,258]. This suggests there may be a consistent trend but yet unproven due to potentially underpowered studies[243]. For NSAID use only the top quintile showed a significant increase in the risk of revision but rates of revision suggest there is a linear trend between amount of the

drug used and subsequent revision risk. However, both the National Institute of Health and Care Excellence and Osteoarthritis Research Society International guidelines advocate using NSAID at the lowest therapeutic dose and avoid long-term use[339,568]. Furthermore, NSAID are recommended to only be used for acute flares[271]. Therefore, higher doses and prolonged use would indicate severity of pain and be a potential indication of an underlying problem with the device. Very few studies have investigated analgesic use and revision rates in THA surgery but those analysing opioid use have found a significant association[243,258].

Using prescription data offers a potentially viable source for post-market surveillance to detect signals of new implants with poorer outcomes compared to proven devices[243,567,569]. In this chapter, patient's prescription data was obtained and evaluated to determine whether the levels of NSAID use or the use of Tramadol in the year following primary THA surgery was associated with subsequent risk of revision. Dispensing of NSAID or Tramadol is potentially a viable alternative for monitoring a patient's pain and therefore, may potentially identify those patients with a higher risk of revision surgery.

One interesting finding from this study is the reduced mortality associated with taking NSAID. A significant linear trend was demonstrated with increasing use of NSAID having higher reduction of mortality risk. There is limited literature investigating NSAID use and mortality with the majority of studies focusing on specific adverse effect as outcome measures[570,571]. NSAIDs are known to have a number of adverse effects on the cardiovascular, gastrointestinal and renal systems[572]. The majority of the

cohort in this chapter are diagnosed with Osteoarthritis. Patients with lower limb Osteoarthritis have been shown to be at increased risk of all-cause mortality[571], and a systematic review has confirmed similar findings for patients with Osteoarthritis in general[573]. The reasons for this increased risk is unknown but may be linked to reduced physical activity levels, especially for patients with Osteoarthritis of the lower limb, and the presence of co-morbidities[574–576]. One study has investigated NSAID use and mortality in patients suffering with Osteoarthritis in nearly 500,000 patients[570]. This study confirmed the finding of this chapter that NSAID use is associated with reduced all-cause mortality in a cohort of Osteoarthritic patients[570]. However, the reasons for this remain unclear and open to speculation[570]. One study has suggested pain is the driver behind increased mortality in Osteoarthritic patients[577], suggesting that perhaps reducing pain may reduce risk of mortality. The study in this chapter found a reduced risk of mortality even when adjusting for co-morbidities which have been shown to be factors of increased mortality in Osteoarthritic patients such as age, gender, diabetes, cancer and cardiovascular disease[574]. Further research is needed to establish the reasons for this observation and whether it is valid in other cohorts.

Medications commonly used to treat pain in such patients include NSAID and opioids[259,578,579]. Classical musculoskeletal complaints do not incorporate neuropathic-like symptoms but there are pathophysiological changes occurring in the central nervous system for patients suffering with chronic musculoskeletal pain[580]. Neuropathic pain is more common than previously thought and is present between 28% to 45% of patients suffering from Osteoarthritis[581–583]. In addition, it is thought the majority of patients suffering with post-operative pain following joint replacement can be treated with neuropathic analgesics[251]. Furthermore, the use of antidepressants for

musculoskeletal pain is significant with up to 3% of all prescriptions used for this indication[299].

Up to a third of patients suffering from hip pain due to Osteoarthritis can present with neuropathic pain[285,584]. Prolonged post-operative pain following THA surgery, which could be an early indicator of failing implants, has been linked to increased rates of neuropathic analgesics use in these patients[251,294]. There are potential significant side effects and reduced analgesic effects following prolonged use of opioid medication offering the opportunity for atypical analgesic use[568,585]. As a result, the use of neuropathic analgesic medication to treat hip pain in the older population has doubled[578].

In this cohort, atypical analgesic use was 12.9% pre-operatively. This was higher compared to other studies with 3.8% antidepressant and 2.7% anticonvulsant use pre-operatively[579]. Post-operative atypical analgesic use in this cohort had fallen to 5.3%. Again, this is higher compared to other studies showing 2.3% of new antidepressant users and 1.4% of new anticonvulsant users, with a total atypical analgesic use of 3.7%, post-operatively in THA surgery[579].

While there may be an increase of analgesic use in the first 4 months with subsequent decrease by the end of the first post-operative year, this was not the case for atypical analgesics such as antidepressants or anticonvulsants[579]. This study was unable to investigate trends of atypical analgesic use within the one-year post-operative period.

However, there was a decrease in total atypical analgesic use post-operatively of approximately 30% which is in keeping with other studies where it has been demonstrated the majority of patients have a reduction in analgesic consumption[236,243,258,579]. Therefore, prolonged use could be an indication of ongoing problems with the device and be a possible early indicator of failure.

Measuring atypical analgesic use offers a potentially valuable surrogate marker of failure for post-market surveillance of THA devices. There are recommendations advocating their use in the treatment of neuropathic like pain in Osteoarthritis[301,305,306,310]. These medications are also only available from a doctor, allowing for a more accurate measurement of use through readily available prescription data. Furthermore, there is evidence of potential addiction if patients use opioid medication for longer than 90 days post-operatively[586–588]. Thus, doctors prescribing medication to treat patients with persistent post-operative pain may favour atypical analgesics.

Healthcare utilisation

Current post-market monitoring is limited as it measures revision rate for any new devices, which has been unable to identify failing implants early[46]. Electronic medical records (EMR) can provide a depth and breadth of information in real-time that may aid in monitoring of new devices. This study has demonstrated that crude GP visit frequency and primary care referrals to the Orthopaedic or Physiotherapy department are associated with increased risk of revision following THA surgery. The potential for

using healthcare utilisation data from EMR systems has been investigated previously[319,325]. Inacio et al showed the frequency and type of post-operative healthcare utilisation was associated with revision in anterior cruciate ligament surgery[325]. This study suggested monitoring frequency of Orthopaedic visits could be an indicator of short-term failure. Prentice et al demonstrated the frequency of post-operative outpatient visits to various departments, including primary care, Orthopaedic and Physiotherapy, are associated with revision rate following THA surgery[319]. This study suggested monitoring healthcare utilisation could be used as a screening tool to identify those patients that may require closer monitoring. In addition, studies have demonstrated other routinely collected data, such as post-operative opioid use, is associated with subsequent revision rate in THA surgery and knee replacement[243,258,589].

Using routinely collected data offers a potentially cost-effective and novel method to identify failing devices early. EMR systems collect all types of patient contact and is not limited to data related to surgery, as is the case in registries[319]. The studies of this chapter investigated encounter types that are potentially early indicators identifying high-risk patients that subsequently require enhanced surveillance. Thus, large primary care datasets, such as the UK Clinical Practice Research Datalink (CPRD)[590], can provide an alternative method of post-operative monitoring in large populations that have undergone THA surgery. The potential of using CPRD data in the evaluation of medical devices is being realised[591]. The annual number of THA replacements performed are increasing across the world presenting a potential challenge in post-operative, prospective monitoring in such large cohorts[319]. Alternative outcome measures, for example patient reported outcome measures, are potentially prohibitively costly[246,253]. In stark contrast, all-cause GP visit rate and referrals to

Orthopaedic or Physiotherapy departments are readily identifiable using routinely collected data in EMR systems. These are potentially cost-effective indicators of post-operative outcome in THA surgery. Exploring the utilisation rates of other healthcare encounter types for patients following joint replacement such as accident and emergency visits has been used as a measure of quality[592–594]. However, EMR systems are able to capture a wider range of patient healthcare encounters not limited to those in secondary care. The healthcare encounters investigated in the studies of this chapter are potential early signals to identify patients who are at higher risk of failure requiring enhanced surveillance or investigations. Comprehensive EMR systems may offer a viable support or alternative for post-market surveillance especially in countries with national databases such as the UK CPRD[590].

Guidelines in the management of Osteoarthritis recommend non-pharmacological and/or pharmacological interventions before the consideration of surgery[260,339,595]. Up to half of patients referred to secondary care will require hip replacement within one year[335]. NICE recommend referral only for those patients with severe or prolonged pain[339]. Routine follow-up for patients who have undergone THA surgery is six to eight weeks and then yearly for novel devices and less frequently for proven implants[341]. Therefore, the expectation for patients post-operatively from the surgeon, GP and patient is only one contact with the Orthopaedic surgeon within the first year. Thus, referral to the Orthopaedic department in the first post-operative year may be a signal for subsequent surgery. In routine practice, it is likely that proposed management by a GP is in agreement with the patient[449] indicating a potentially significant post-operative issue if referral to the Orthopaedic department is required.

The study here has confirmed an association between Orthopaedic referral and subsequent revision. There are recent recommendations for reduced contact with the Orthopaedic department using virtual clinics[596]. Thus, using referral to the Orthopaedic department for surveillance in the future may be a stronger indicator of a significant problem than highlighted here.

Physiotherapy has been available since the 1970's for primary care physicians to directly refer to in both the community or secondary care settings[597]. NICE guidance recommends patients with joint pain should be offered non-operative treatments, such as Physiotherapy, before considering surgery[339]. Patients receiving Physiotherapy have been demonstrated to increase the risk of subsequent Orthopaedic referral for consideration of joint replacement[449], confirming primary care physicians follow guidelines by exhausting all non-operative options prior to referral. In the cohort here there were more referrals to the Physiotherapy department compared to the Orthopaedic department suggesting primary care physicians may see it as a management option prior to surgery. Both the British Orthopaedic Association and NICE recommend Physiotherapy input during the immediate post-operative period until discharge but no formal interaction thereafter[341,342]. As with Orthopaedic referrals, referrals to the Physiotherapy department following THA surgery is likely to be representative of a deviation from normal post-operative recovery.

Significant and prolonged pain has been identified as one of the most important factors in determining whether to undergo total joint replacement[335,338,598]. As the patient is not routinely followed-up within the first post-operative year after the eight-week mark, and primary care physicians attempt to manage patients in the community, monitoring

healthcare utilisation is likely a marker of symptomatic post-operative issues such as pain. Moreover, in the UK there are up to 1,000,000 annual Orthopaedic outpatient appointments[599] with significant breaches of approximately 8,000 per month[600]. Follow-up is designed to detect failing implants or post-operative complications[599]. However, due to the vastly low ten year revision rate and the failings to meet waiting list targets, a diversity of follow-up protocols across the UK have emerged[599] with some discharging patients following the initial post-operative appointment[601]. Therefore, monitoring referral rates to the Orthopaedic or Physiotherapy department would offer a useful additional early screening tool to identify patients at risk of device failure, reduce the need for additional follow-up appointments and subsequently reduce the pressures of current follow-up requirements. Higher healthcare utilisation has been shown to be more predictive of subsequent surgery compared to patient or disease characteristics[449]. Furthermore, referrals to either the Orthopaedic or Physiotherapy departments are a potentially good measure of significant post-operative problems as GPs will only refer after assessing the severity of the presenting complaint and the significant impact to the patient[330]. Therefore, monitoring healthcare utilisation may be an excellent early signal for potentially failing implants and aid in post-market surveillance in large cohorts. Healthcare utilisation monitoring could be used in other countries that have a similar process whereby GPs act as the gate keepers to secondary care such as Canada[416], Denmark[343], and the Netherlands[401,602]. In such healthcare systems, the majority of healthcare utilisation for musculoskeletal problems are those patients visiting their primary care physician[416].

3.5.1. Limitations

General limitations will be addressed in Chapter 5, and specific limitations to the studies of this chapter are mentioned here.

Compared to the sample size calculation the cohort did not have enough power to draw formal conclusions regarding bearing stratification due to missing bearing information and a subsequent complete case analysis. Therefore, further research, would be needed to investigate whether the exposures of this thesis can detect the poor outcome of MoM implants early.

Analgesic utilisation

One limitation is the fact that a record of an analgesic prescription does not necessarily equate to the patient ingesting the medication. It is challenging to accurately measure compliance in observational studies[258], especially in the context of binary exposure variables. While analysing prescription medication is a good objective measure of analgesic use, a prescription does not necessarily translate to use as full adherence may only be achieved in 50% of patients[261].

Paracetamol and codeine phosphate use was not analysed here due to insufficient data. Paracetamol is available over the counter in most countries and rarely prescribed, thus data on its utilisation is difficult to obtain. Codeine phosphate is a commonly used opioid medication but is also available over the counter. However, its use is not recommended for patients with musculoskeletal pain[260]. It is possible patients who are

not requiring or cannot take regular NSAID or Tramadol medication for pain are instead using other analgesics. Thus, patients falling within the first quintile of NSAID use or counted as non-Tramadol users, who have not required revision surgery, may be taking other analgesic medications.

The reasons for the prescription are not found in the Catalan linked database and is often poorly recorded in primary care records[603]. All prescription data obtained was assumed to be for the treatment of pain related to the hip involved with the primary THA surgery. There are multiple reasons a medication can be prescribed. It is possible that the analgesics prescribed were for the treatment of pain conditions rather than specifically for the hip that was operated on. Furthermore, there may be different prescribing practices and each surgeon could have preferences to medication used. This data is unavailable here and was not accounted for in the statistical model, and thus could lead to potential confounding. In addition, it is conceivable that patients received opioid or NSAID medication outside the healthcare system captured in the Catalan databases and thus not included here. It is possible these patients were prescribed atypical analgesics for other reasons such as neuropathic pain originating from another location other than the hip or for the treatment of neuropathic pain of other causes such as diabetes or cancer. Furthermore, patients with pre-operative analgesic use has been shown to increase the risk of subsequent post-operative use[579]. Thus, patients using atypical analgesics pre-operatively could be at higher risk of persistent post-operative use[579]. However, despite these potential issues, nearly 40% of patients using antidepressants pre-operatively saw either a decrease or cessation of use post-operatively[579]. In this study, pre-operative users of atypical analgesics were

excluded to minimise potential confounding effects. In addition, multiple confounders, including history of cancer and presence of diabetes, were adjusted for in the final regression analysis.

Pain is not exclusively managed by pharmaceutical means and patients may choose to manage their hip pain through other conservative measures. However, this is unlikely to affect final estimates towards the null hypothesis[604] and therefore the results here are likely to represent conservative estimates of risk of revision.

For all analgesic use this chapter recognises that there are multiple methods of evaluating the exposure such as cumulative quantity measured as daily defined dose, average daily use as well as binary. Unfortunately, the dataset did not offer to analyse the Tramadol exposure in any other way than binary which could provide more information and richer analysis with potential cut-offs. Thus, the results here should recognise this when comparing to other studies. However, these studies are exploratory in nature to examine for an association of analgesic use to revision risk.

NSAID are readily available over the counter and approximately one fourth[266] to one third[605] of all NSAID sold are bought in this way. A significant number, over a quarter, of patients do not redeem prescriptions of analgesic medication prior to surgery which may be accounted for by over the counter drug use[266]. However, the proportion of patients using over the counter medication is likely to be much smaller for those with "chronic painful conditions", such as osteoarthritis[266]. Furthermore, patients with

severe pain are more likely to use prescribed medications[261]. Moderate to severe pain would indicate significant issues post-operatively and would indicate review with consideration for revision surgery. Furthermore, moderate to severe pain is associated with higher risk of revision surgery[255]. It is unlikely over the counter medication use will sufficiently bias association between prescription analgesic use and risk of revision.

Pre-operative opioid use has been demonstrated as a risk factor for subsequent prolonged post-operative use[236], particularly for those using antidepressants or benzodiazepines[588]. While these specific medications are not associated with subsequent revision, depression and anxiety were adjusted for in the survival model which would capture patients on these medications. Intensive pre-operative users of opioids may have a significant effect on patients using these medications post-opertively[270]. However, the overall proportion of such users was small in their study. In the RACat cohort, it is not possible to ascertain the proportion of intensive opioid users pre-operatively.

There have been studies suggesting that NSAID use in the peri or post-operative period can cause reduced osseointegration and subsequent implant failure, particularly in uncemented devices that rely on osseointegration for stability[606–610]. In this regard NSAID use could not be used as a surrogate measure of failure. The expectation would be that those patients who have had cementless fixation and taking regular NSAID medication would be at risk of early failure. However, this study showed there to be no significant interaction with cement use. The results were stratified to show failure

compared to cemented, hybrid/reverse and cementless fixation: no significant difference was found. This correlates with literature findings that NSAID use linked to reduced osseointegration is inconclusive in clinical studies. Furthermore, NSAIDs are routinely used in the prevention of Heterotrophic Bone Formation (HBF) and studies have found no increase in aseptic loosening[611,612]. In a study performed on the Danish Hip Arthroplasty register, NSAID use for HBF had no increased risk of revision in cementless implants and, reduced the revision risk in cemented implants[613].

One study found an increased risk of revision associated with NSAID use but concluded patients were likely using NSAID as an analgesic. Therefore, NSAID would be considered as an indication of "poorly functioning implant" rather than an independent risk factor[265], supporting this study that NSAID use can potentially be used as a surrogate maker of implant survival.

Due to the relatively low use of atypical analgesic medication for musculoskeletal pain it is possible that preferences of patients or physicians could influence trends of use even during the study period. Furthermore, marketing trends or changes in regulation of medications can influence prescription[578]. Both Pregabalin and Gabapentin have been classed as controlled drugs recently and concerns regarding potential addiction could have influenced rates of prescribing[614]. Unfortunately, this cohort lacks sufficient information to address these potential confounders.

Healthcare utilisation

There is a lack of information on reasons for GP encounters, thus it may be possible patients are not attending due to hip pain from their THA. Patients may visit their

primary care physician for more than one issue and any discussion concerning their THA surgery may be incidental. These studies lack granular details in regard to specific reasons for each encounter. The information gathered is limited to the SIDIAP classification for encounter information, which could potentially limit the findings of this study to those EMR systems that have common metadata definitions[325].

It is possible the SIDIAP primary care dataset has missing or misclassified visit type and therefore, these would not have been included in the studies conducted here. Missing data could potentially bias results. However, the aim of these studies was to ascertain whether raw primary care visit frequency is associated with revision rather than constructing a predictive model.

Information on indications for referral to the Orthopaedic or Physiotherapy department was not available. It is possible that referrals to either may have been part of a prescribed management plan specific to the patient but due to administrative errors, the patient was not booked by the hospital subsequently requiring a GP referral. The majority of patients are usually discharged by the Physiotherapist at the time of discharge from the hospital or within one week. British Orthopaedic Association guidelines also recommend Physiotherapy input during the immediate post-operative phase and ensure discharge when the patient is "considered capable of coping in the environment at their destination"[341]. It is likely there are no expected routine Physiotherapy follow-up appointments. In addition, there is usually only one expected Orthopaedic follow-up within the first post-operative visit and then at the one-year mark[341,342]. Patients in these studies were not included if referred within eight weeks.

Therefore, any additional referral from the GP within the time period of this study is likely to represent an unexpected post-operative problem.

For the referrals to the Orthopaedic and Physiotherapy department exposures, patients with referrals within eight weeks post-operatively were excluded. This was based on gold standard post-operative follow-up according to NICE and BOA guidelines[341,345]. However, unlike the UK, Spain has a decentralised healthcare system with no centralised guideline such as NICE, resulting in potentially differing guidelines between the two countries[378] It is possible a significant number of hospitals and surgeons have varying follow-up protocols and therefore the thresholds used to exclude patients in this study could potentially be incorrect. For example, referrals made within the first post-operative year in Spain could represent patients who would have been routinely followed up following THA surgery but were missed due to administrative error. However, over 95% of patients in Spain undergoing primary THA surgery are discharged home and both patients and doctors ensure discharge does not occur until "the patient has recovered a certain degree of autonomy"[553]. Therefore, Spanish Physiotherapy referrals are likely to reflect similar practices to the UK, but Orthopaedic follow-up may still differ between the two countries. Despite no centralised guidelines, it is unlikely that Orthopaedic follow-up deviates considerably from UK guidance and Spanish literature cites the UK surveillance system as a gold standard example[350]. Furthermore, the cohort here has significant coverage of the Catalan region capturing differing Orthopaedic practices.

There are a wide variety of reasons for referral to the Orthopaedic or Physiotherapy department by the primary care physician. However, some of these reasons may not be

indicative of a failing device. For example, primary care physician burnout has been shown to increase rates of referral to secondary care as objective workload increases[615]. General workload in primary care has substantially increased including higher consultation rates and longer consultation durations during the period of this study[616]. Furthermore, patients who have a belief that their disease will leave a permanent pain or disability are likely to have increased healthcare utilisation[330].

It has been suggested there are deficiencies for managing musculoskeletal conditions, with particular regard to arthritis, in primary care due to a lack of confidence[617,618]. Therefore, GPs may have gaps in knowledge of the management of musculoskeletal conditions leading to referrals for diagnostic and treatment purposes[401]. These factors have the potential to lead to poor or inappropriate specialist referrals[416,619]. The studies in this chapter are unable to evaluate the quality of referral and whether it was accepted or rejected by the Orthopaedic or Physiotherapy department.

Characteristics of a GP surgery and the GP have been shown to influence rates of referral to the Physiotherapy department[343]. Increased rates were found to be associated for city practices, and male GPs[343]. Despite no specific data regarding such characteristics, the SIDIAP database has excellent coverage of GP surgeries in the Catalan region.

Rates of referral to the Physiotherapy department are significantly associated and dependant on available resources, such as waiting lists or access to onsite services, and

individual Physiotherapist's ability to promote their work[620,621]. Interestingly, these factors may be irrespective of local or national guidelines[622]. No data on Physiotherapy availability is held in the RACat or SIDIAP databases and could leave residual confounding.

Qualitative research has suggested other influences of rates of Physiotherapy referral include higher rates for patients perceived as "heartsink" to alleviate the pressure such patients can inflict on workload[622]. Even if the primary care physician feels the patient would not benefit from Physiotherapy a referral is still made to relieve the GP's own stress[622]. Furthermore, a primary care physicians future behaviour with Physiotherapy referrals can be dependent on reports received from previous referrals or past exeperience[622]. These could represent residual confounding.

3.5.2. Strengths

General strengths will be addressed in Chapter 5, and specific strengths to the studies of this chapter are mentioned here. Aside from death, which was accounted for in competing risk regression, no attrition existed and therefore all events for prescription data were captured. Using prescription data ensured no patient response bias occurred via lack of response or poorly completed forms due to incomprehension[623] offering a more accurate and complete assessment of pain.

Ultimately, the intention of these studies is to investigate whether all-cause healthcare resource utilisation, with routinely collected data, is associated with revision risk, irrespective of the nature of care. This data is the most readily available and easily

extractable from an EMR system, which would allow for real-time monitoring, expedited screening and potential further investigation or intervention for those patients deemed high risk of failure.

The frequency of GP visits is potentially generalisable to a UK cohort. A large study has demonstrated the average frequency of GP visits per patient per year during a similar period to that of this study (2007-2014) in England was approximately four[616].

The Spanish healthcare system is part of a national health service and access to secondary care is only through primary care[350]. Therefore, the results of these studies are potentially generalisable to similar healthcare systems but may not apply to those with a predominantly private system, or those that allow direct access to secondary care.

3.6. Conclusion

The importance of post-market surveillance with observational studies using linked registry data has been highlighted in this chapter. The RACat is linked to primary care data and offers a unique opportunity to explore potential surrogate markers of THA surgery for use in post-market surveillance of devices newly introduced to the market. This chapter investigated novel potential surrogate markers of subsequent revision in patients who underwent primary THA. Analgesic and healthcare resource use that are routinely measured were found to be associated with subsequent revision rate following primary THA surgery. This suggests that analgesic and healthcare use can potentially be used as a surrogate measure for the early detection of patients at increased risk of failing devices.

4. Investigation of Novel Surrogate Markers of Total Hip Arthroplasty Outcome to Detect Differences between Device Bearings

4.1. Summary

Chapter 3 highlighted the potential for analgesic and healthcare utilisation to be used as a screening tool for early identification of failing THA devices. This chapter investigates the revision rate of metal-on-metal bearings within the RACat cohort against other bearings. This chapter further investigates the use of novel surrogate markers to detect differences between metal-on-metal and other bearings.

4.2. Introduction

Monitoring alternative outcomes, such as healthcare utilisation, has been suggested as a screening tool to identify devices that may require closer monitoring[319]. Chapter 3 demonstrated an association between analgesic or healthcare utilisation within the first post-operative year and subsequent revision rate in Total Hip Arthroplasty (THA). Metal-on-Metal bearings have gained notoriety due to higher-than-expected revision rates, with many brands being recalled from the market[205]. However, it is unclear if monitoring analgesic or healthcare utilisation as outcomes post-operatively will detect differences in revision rates between bearing devices.

The main aim of this experimental chapter is to determine whether analgesic or healthcare utilisation could be used as an outcome measure to detect differences in revision rate of different bearing devices.

4.3. Methods

This chapter builds on the results from the previous chapter. Further details of the data source, ethics approval, study population and multiple imputation can be found in chapter 3.

4.3.1. Aims

To investigate the numbers needed to screen to detect failing implants using novel surrogate markers.

To investigate the differences in revision rate between bearing devices in primary Total Hip Arthroplasty.

To investigate the differences in analgesic or healthcare utilisation between bearing devices in primary Total Hip Arthroplasty.

4.3.2. Study design

Cohort studies.

4.3.3. Outcome

The first outcome was all-cause revision hip arthroplasty from the first post-operative year up to 7.5 years following primary THA surgery. Revision was defined as any exchange, removal, or addition of a prosthesis component of the primary prosthesis.

The second outcome was analgesic or healthcare utilisation, as defined by the exposures in chapter 3, measured within the first post-operative year following primary THA surgery. Outcomes used can be stratified into two categories:

- Analgesic use within the first year post-operatively:
    - Non-Steroidal Anti-Inflammatory Drugs (NSAID)
    - Tramadol
    - Atypical Analgesics (Pregabalin, Gabapentin, Amitriptyline)
- Healthcare resource use within the first year post-operatively:
    - General Practice visits
    - Referrals to the Orthopaedic department
    - Referrals to the Physiotherapy department

Tramadol, atypical analgesics, and referrals to the Orthopaedic or Physiotherapy department were binary variables as described in chapter 3. Following analysis in chapter 3, only the top quintile of NSAID use was found to be significantly associated with revision rate, and more than four primary care visits was found to be clinically significant[377]. Therefore, both were transformed to binary variables using the top quintile versus other use, and more than four primary care visits versus less than four.

4.3.4. Exposures

The granularity of bearing devices used for primary THA surgery in this cohort include, Metal-on-Metal (MoM), Metal-on-Poly (MoP), Ceramic-on-Ceramic (CoC), and Ceramic-on-Poly (CoP). The majority of bearing devices used was MoP (43.6%). The bearing of interest for this experimental chapter is MoM due to the recent increased revision rates and recall of devices[205]. In the latest National Joint Registry report, MoP, CoC and CoP have similar revision rates with a clear difference to MoM devices[624]. Therefore, the exposure used was a binary variable comparing MoM bearings to all other bearings.

4.3.5. Confounders

A priori, variables were identified as potential confounders in relation to failure based on clinical knowledge, literature search, and biological plausibility. Measurements were taken at the time of or prior to surgery as registered by the SIDIAP or RACat databases.

The potential confounding variables identified in the RACat and SIDIAP joint database included: Age, Sex, Socio-economic status, Charlson Co-Morbidity Index, Depression, Atrial Fibrillation, Hypertension, Cardiovascular Disease, Diabetes, Cancer, Indication for Surgery, Previous Fixation, Previous Fracture, Body Mass Index (BMI), Smoking status, Alcohol status. A detailed literature review of these confounders and their association with exposures and revision in primary THA surgery can be found in Chapter 3.

4.3.6. Statistical Analysis

Numbers needed to teat for survival analysis can be calculated using the difference in survival function at a specified time[625]. Survival function at the maximum follow up time was used for each of the novel surrogate markers to estimate numbers need to screen to identify one revision.

Descriptive statistics was used to describe the cohort sample and exposures. For the three confounding variables with substantial proportions of missing data (up to 40%), multiple imputation by chained equations was applied to ensure analysis can be performed on a complete cohort to reduce the risk of selection bias and increase precision from the studies being otherwise potentially underpowered. Subsequent analysis was performed on individual datasets and combined in conjunction with Rubin rules using in built Stata commands. Further details of the multiple imputation technique can be found in chapter 3.

Survival analysis was used to model the association between bearing and revision surgery. Kaplan-Meier estimates of the cumulative incidence of revision surgery stratified by exposure in the first post-operative year were plotted. No significant association was found between bearing devices and mortality. Therefore, Cox proportional hazards model (Cox regression) was used to estimate hazard ratios (HR) and 95% confidence intervals according to exposure without the need to adjust for the competing risk of mortality[550]. Multivariable Cox regression models were used to adjust for pre-identified confounders as previously discussed. A more detailed explanation and argument for multiple imputation methods, and for the use of survival analysis are found in Appendix 2 and 3 respectively.

The novel surrogate markers used as the outcome measure in this study were measured within the first post-operative year. As a result, there are no concerns with differing follow up time and censoring that require complex survival analysis. Therefore, logistic regression was used to estimate the odds ratio (OR) and 95% confidence intervals according to device bearings against surrogate markers of THA outcome[626].

All statistical analysis including multiple imputation was performed using Stata 15 (StataCorp. 2017)[542] statistical software.

Sample size

A sample size calculation was performed in relation to detect differences in revision rate between MoM and other bearings. A detailed description can be found in chapter 3.

To detect differences in surrogate marker rates a sample size calculation was performed using the "power cmh" function in Stata[627]. This calculates a minimum sample size for a Cochran–Mantel–Haenszel test of association in stratified two by two tables. Sub-hazard ratios determined in chapter 3 was used to approximate effect size, and 80% power at 5% level of significance (alpha) (p-value of 0.05) was adopted.

4.4. Results

Results general to this cohort such as baseline characteristics, participant follow up and lifetables can be found in chapter 3.

4.4.1. Study population and baseline characteristics

Data from the RACat registry was found for 17,034 who underwent primary Total Hip Arthroplasty (THA) surgery from 2005 to July 2012. Of these 270 (1.6%) had revision surgery within the first post-operative year, 331 (1.9%) were under the age of 40 years, and 6,997 (41.1%) had no documentation of the bearing used. These patients were excluded leaving our study population as 9,436. From those patients included, 534 (5.7%) underwent THA surgery using a metal-on-metal bearing device, and 8,902 (94.3%) used other bearing devices including 7,160 metal-on-poly, 606 ceramic-on-ceramic, and 1,136 ceramic-on-poly bearing devices (Figure 4-1).

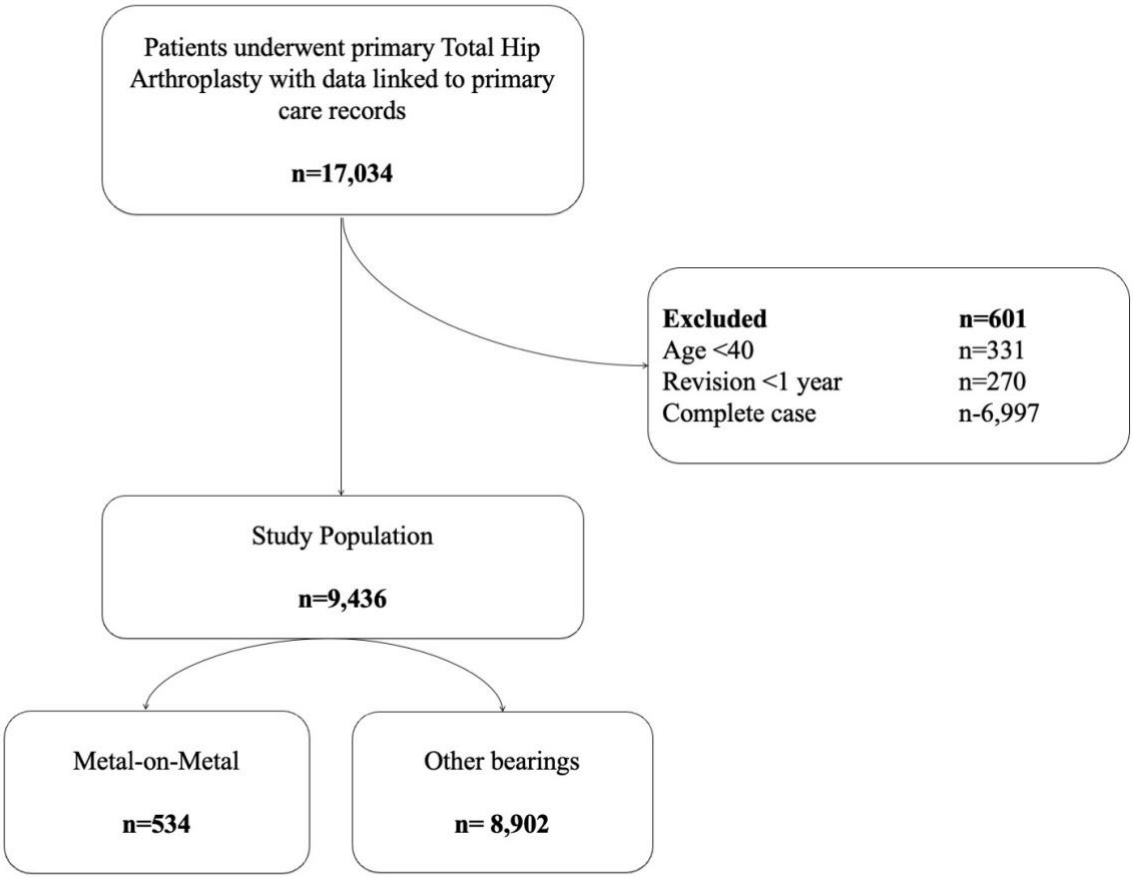


Figure 4-1: Metal-on-Metal bearing analysis patient selection flow-chart.

Baseline characteristics for the study population following multiple imputation can be found in Table 4-1.

Table 4-1: Baseline characteristics for THA bearings.

		Other bearings	MoM bearing
Characteristic			
BMI	Observation Mean (SD)	29.2 (±4.8)	29.7 (±4.6)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**
Age	40 to <60	1,309 (13.9)	361 (3.8)
	60 to <70	2,147 (22.8)	159 (1.7)
	70 to <80	3,598 (38.1)	12 (0.1)
	80 and older	1,848 (19.6)	2 (0.02)
Gender	Male	3,922 (41.6)	427(4.5)
	Female	4,980 (52.8)	107 (1.1)
Alcohol	No	6,080 (64.4)	241 (2.6)
	Yes/Ex	2,822 (29.9)	293 (3.1)
Smoking	Never	6,380 (67.6)	252 (2.7)
	Current/Ex	2,522 (26.7)	282 (3.0)
Charlson Score	0	4,575 (48.5)	380 (4.0)
	1	359 (3.8)	26 (0.3)
	≥2	3,968 (42.0)	128 (1.4)
Socio-economic status	1	1,425 (15.1)	76 (0.8)
	2	1,559 (16.5)	97 (1.1)
	3	1,594 (16.9)	76 (0.8)
	4	1,588 (16.8)	99 (1.1)
	5	1,517 (16.1)	107 (1.1)
	Rural	1,219 (12.9)	79 (0.8)
CVD	No	8,335 (88.3)	517 (5.5)
	Yes	567 (6.0)	17 (0.2)
AF	No	8,440 (89.4)	526 (5.6)
	Yes	462 (4.9)	8 (0.1)
Hypertension	No	4,033 (42.7)	386 (4.1)
	Yes	4,869 (51.6)	148 (1.6)
Diabetes	No	7,354 (77.9)	491 (5.2)
	Yes	1,548 (16.4)	43 (0.5)
Depression/Anxiety	No	7,259 (76.9)	463 (4.9)
	Yes	1,643 (17.4)	71 (0.8)
Cancer	No	8,567 (90.8)	524 (5.5)
	Yes	335 (3.6)	10 (0.1)
Indication	Fracture	1,260 (13.4)	8 (0.1)
	Osteoarthritis	6,943 (73.6)	470 (4.9)
	Other	699 (7.4)	56 (0.6)
Previous Fixation	No	8,359 (88.6)	497 (5.2)
	Yes	543 (5.8)	37 (0.4)
Previous Fracture	No	7,930 (84.0)	513 (5.5)
	Yes	972 (10.3)	21 (0.2)

Characteristics of THA bearing against novel surrogate markers for the study population following multiple imputation can be found in Table 4-2.

Table 4-2: Characteristics for THA bearings against novel surrogate markers

Surrogate Marker		Other bearings (%)	MoM bearing (%)
NSAIDs	Other	6,911 (73.2)	435 (4.6)
	Top quintile	1,991 (21.1)	99 (1.1)
Tramadol	Non-users	6,955 (73.7)	439 (4.7)
	Users	1,947 (20.6)	95 (1.0)
Atypicals	Non-users	8,375 (88.8)	511(5.4)
	Users	527 (5.6)	23 (0.2)
GP Visits	≤4	3,205 (34.0)	191 (2.0)
	>4	5,697 (60.4)	343 (3.6)
Ortho referral	Non-referral	8,230 (87.2)	494 (5.3)
	Referral	672 (7.1)	40 (0.4)
Physio referral	Non-referral	8,134 (86.2)	474 (5.0)
	Referral	768 (8.1)	60 (0.7)

4.4.2. Numbers needed to screen

The survival functions for the novel exposures at the maximum follow up time can be found in Table 4-3.

Table 4-3: Survivor function for each surrogate marker at the maximum follow up time interval.

Surrogate Marker	Category	Time Interval Years	Survivor Function (95% CI)
NSAIDs	First Quintile (p<20)	7.4	0.9904 (0.9684 - 0.9971)
	Fifth Quintile (p≥80)	7.4	0.9540 (0.9136 - 0.9758)
Tramadol	Non-users	7.4	0.9510 (0.8702 - 0.9820)
	Users	7.4	0.9678 (0.9504 - 0.9792)
Atypical analgesics	Non-users	6.8	0.9538 (0.8958 - 0.9799)
	Users	6.8	0.9714 (0.9354 - 0.9875)
GP visits	0-4 visits	7.4	0.9935 (0.9823 - 0.9976)
	>4 visits	7.4	0.9477 (0.8787 - 0.9780)
Ortho referral	Non-referral	5.2	0.9795 (0.9679 - 0.9869)
	Ortho referral	5.2	0.9596 (0.9211 - 0.9795)
Physio referral	Non-referral	5.3	0.9768 (0.9629 - 0.9855)
	Physio referral	5.3	0.9193 (0.7231 - 0.9784)

The estimated numbers needed to screen for each surrogate marker can be found in Table 4-4.

Table 4-4: Number needed to screen for each surrogate marker

Surrogate Marker	Number Needed to Screen
NSAIDs	28
Tramadol	60
Atypical analgesics	57
GP visits	22
Ortho referrals	51
Physio referrals	18

4.4.3. Rates of revision

MoM bearing was associated with an increased incidence of revision, 7.04/1000 (3.79-13.08) compared to other bearings with 3.47/1000 (2.81-4.29) (Table 4-5). The median (inter-quartile range) of time to revision is shown in Table 4-5. This finding is confirmed in the Kaplan-Meier estimates of cumulative risk of revision. (Figure 4-2).

Table 4-5: Unadjusted rates of revision for MoM versus other bearing devices.

Bearing	Number (%)	Number of Revisions	Median (inter-quartile range) of years to revision	Revision rate /1,000 person-years (95% CI)
Other	8,902 (94.3)	86	1.9 (1.5-3.0)	3.47 (2.81-4.29)
MoM	534 (5.7)	10	1.6 (1.3-3.1)	7.04 (3.79-13.08)

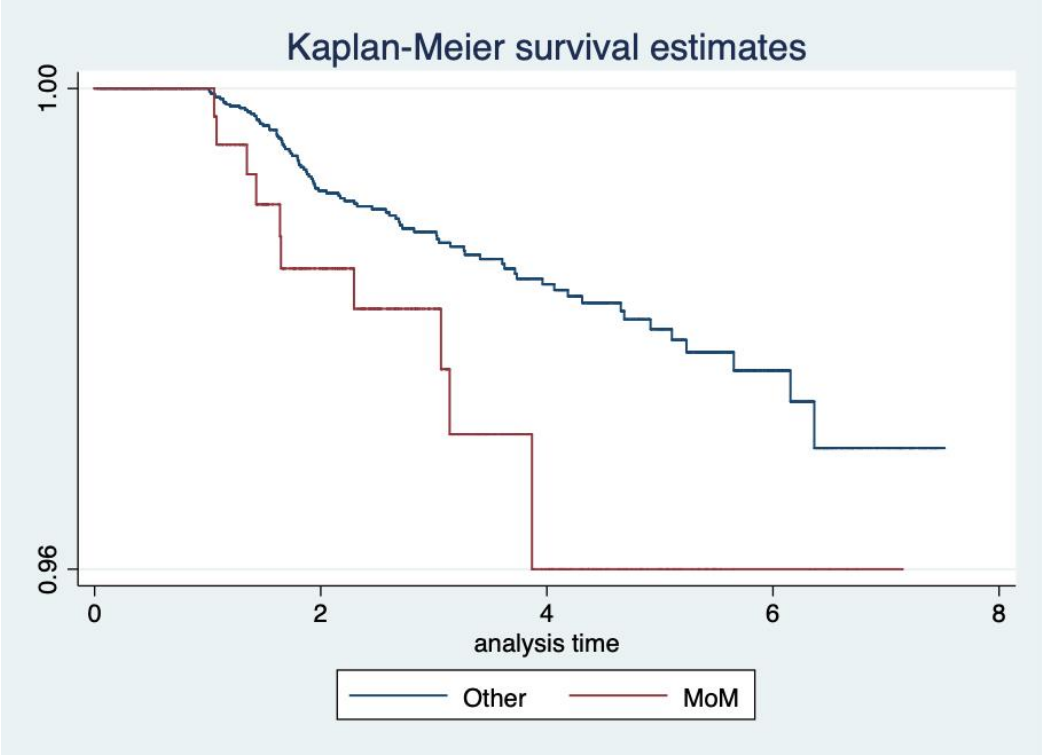


Figure 4-2: Kaplan-Meier survival estimates for device bearing.

4.4.4. Device bearing versus mortality

There was no statistically significant association between device bearing and mortality within the cohort, suggesting that mortality would not need to be adjusted for in the regression using Fine & Gray modelling (Table 4-6).

Table 4-6: Univariate and multivariate Cox proportional hazards regression for device bearing against mortality.

	Univariate HR (95% CI)	**Multivariable HR (95% CI)**
Bearing vs mortality	0.24 (0.12-0.52)	0.70 (0.32-1.53)
	p=0.000	p=0.365

4.4.5. Device bearing versus revision

Survival analysis demonstrated rates of revision after the first-year post operation were positively associated with metal-on-metal bearings. However, this finding was not statistically significant.

THA participants who received a metal-on-metal bearing device compared to those receiving other device bearings demonstrated an adjusted Hazard Ratio (HR) of 1.68 [0.81-3.49; p=0.163] (Table 4-7).

Table 4-7: Univariate and multivariate Cox proportional regression analysis to investigate the association with device bearing and subsequent revision rate.

	Univariate HR (95% CI)	**Multivariable HR (95% CI)**
MoM bearing vs. Other bearings	1.99 (1.03-3.84)	1.68 (0.81-3.49)
	p=0.039	p=0.163

4.4.6. Sample size

Estimated sample size for each analgesic and healthcare utilisation surrogate marker can be found in Table 4-8.

Table 4-8: Estimated sample size for each analgesic and healthcare utilisation surrogate marker.

	N total required (per stratum)	N patients in smallest category
NSAID	470 (235)	1,991
Tramadol	374 (187)	1,947
Atypicals	864 (432)	527
GP visits	442 (221)	5,697
Ortho referrals	1,686 (843)	672
Physio referrals	1,100 (550)	768

4.4.7. Device bearing versus novel surrogate markers

Logistic regression demonstrated rates of NSAID use and referral to the physiotherapy department, measured within the first-year post operation, were significantly associated with metal-on-metal bearings. No statistically significant difference between bearing device was demonstrated for the surrogate markers of Tramadol, atypical analgesic, GP visits, and referral to the Orthopaedic department.

THA participants who received a metal-on-metal bearing device compared to those receiving other device bearings demonstrated an adjusted Odds Ratio (OR) of 1.32 [1.03-1.68; p=0.028], when NSAID use was the outcome.

THA participants who received a metal-on-metal bearing device compared to those receiving other device bearings demonstrated an adjusted OR of 1.04 [0.81-1.33; p=0.765], when Tramadol use was the outcome.

THA participants who received a metal-on-metal bearing device compared to those receiving other device bearings demonstrated an adjusted OR of 0.93 [0.59-1.48; p=0.771], when atypical analgesic use was the outcome.

THA participants who received a metal-on-metal bearing device compared to those receiving other device bearings demonstrated an adjusted OR of 1.09 [0.90-1.33; p=0.370], when GP visits was the outcome.

THA participants who received a metal-on-metal bearing device compared to those receiving other device bearings demonstrated an adjusted OR of 0.99 [0.70-1.42; p=0.978], when referral to the Orthopaedic department was the outcome.

THA participants who received a metal-on-metal bearing device compared to those receiving other device bearings demonstrated an adjusted OR of 1.54 [1.13-2.09; p=0.006], when referral to the Physiotherapy department was the outcome. (Table 4-9).

Table 4-9: Univariate and multivariate logistic regression analysis to investigate rates of novel surrogate maker between metal-on-metal bearings versus other bearings.

Surrogate marker as Outcome	Univariate OR (95% CI) MoM bearing vs. Other bearings	Multivariable OR (95% CI) MoM bearing vs. Other bearings
NSAID	0.79 (0.63-0.99) **p=0.039**	1.32 (1.03-1.68) **p=0.028**
Tramadol	0.77 (0.62-0.97) **p=0.026**	1.04 (0.81-1.33) p=0.765
Atypicals	0.70 (0.45-1.07) p=0.100	0.93 (0.59-1.48) p=0.771
GP visits	1.01 (0.84-1.21) p=0.912	1.09 (0.90-1.33) p=0.370
Ortho referral	0.99 (0.71-1.38) p=0.961	0.99 (0.70-1.42) p=0.978
Physio referral	1.34 (1.01-1.77) **p=0.039**	1.54 (1.13-2.09) **p=0.006**

4.5. Discussion

This chapter further investigated the effectiveness of novel surrogate markers in detecting differences between THA device bearings. Rates of Tramadol and atypical analgesic use within the first post-operative year were not significantly different between Metal-on-Metal (MoM) and other bearing devices. Attending more than four visits and being referred to the Orthopaedic department within the first post-operative year demonstrated no significant differences between patients receiving a MoM or other device. However, rates of NSAID use in the top quintile and those patients being referred to the physiotherapy department were significantly different between bearing devices. Those patients who underwent THA surgery with a MoM bearing had approximately a 30% and 50% increased risk of being in the top quintile of NSAID utilisation and referred to physiotherapy within the first post-operative year. This suggests that measuring NSAID use and physiotherapy referral rates post-operatively can potentially be used to identify differences in revision rate between devices.

The numbers needed to screen ranged from 18 for referrals to the Physiotherapy department to 60 for Tramadol use. This suggests 18 to 60 patients, followed up for one year post-operatively, would be needed to detect one patient undergoing revision surgery. Furthermore, the required number needed to screen is comparable to Radiostereometric Analysis (RSA) in which a sample size of 15 to 25 patients per stratum (total of 30 to 50 patients) is required to identify differences between devices within two years post-operation[134,136,157]. The estimated sample sizes required for the novel surrogate markers ranged from 527 to 1,991 for all except GP visits with a sample size of 5,697. This is comparable to the minimum sample size for a non-inferiority study using contemporary benchmarking standards (3,200), and the Kaplan Meier

method (1,600)[125]. The ASR implant had an excellent early ODEP rating of 3A depicting an acceptable failure rate of less than 3% after three-year follow-up data[46]. Therefore, monitoring surrogate markers could detect outliers earlier and with fewer numbers compared to monitoring revision rate. In addition, analgesic and healthcare utilisation uses routinely collected data offering a potential cost-effective screening mechanism.

One of the main limitations in current methods of post-market surveillance is the reliance on revision rate[105]. New devices to the market are typically introduced in small numbers and revision is a rare outcome, thus making failing implants difficult to identify[105]. In the case of the ASR device, single centre cohort studies were able detect its failings prior to the NJR identifying them as outliers[62,65]. Therefore, markers of THA outcome are sought to potentially detect outliers earlier. In this study, there was no significant difference in revision rate between MoM and other bearings during a relatively short median follow-up time which correlates with other studies[348,628]. However, NSAID use and referral to the Physiotherapy department were able to detect a significant difference between MoM and other bearings, suggesting that these novel surrogate markers may be able to detect outliers earlier than revision.

There were no obvious correlations between sub-hazard ratio (SHR) or area under receiver operator curve (AUC). The top quintile of NSAID use and referral to the physiotherapy department were found to have an SHR of 2.20 and 2.07 respectively. However, tramadol and atypical analgesic use had higher SHRs of 2.42 and 2.59 but was unable to detect a significant difference between bearings. Four or more primary care visits and referral to the Orthopaedic department had similar or lower SHRs of 2.10 and 1.90 respectively. Therefore, the strength of association to revision rate may not determine a surrogate marker's ability to detect differences in bearings early.

Furthermore, there were no obvious correlations between specificity and a surrogate marker's ability to detect a significant difference between bearings. Although referral to the physiotherapy department was found to have a 92% specificity, atypical analgesic use and referral to the Orthopaedic department were higher at 94.0% and 93.1% respectively. In contrast, NSAID use had a lower specificity of 80.2% yet was able to detect a significant difference between bearings. Despite Tramadol use having a similar specificity at 80.6%, it was unable to detect any significant difference. Finally, there were no obvious correlations between the area under the curve (AUC) and significant findings in this study. NSAID use was shown to have the highest AUC of 0.6212, however, referral to the physiotherapy department had one of the lowest AUC of 0.5225 but was still able to detect a significant difference between bearings.

The reasons for this are unclear. However, both analgesic and healthcare utilisation was measured within the first post-operative year. Those surrogates finding a significant difference between bearings may have done so because their use early in the management of post-operative pain could be more likely during this period. NSAIDs are considered the "mainstay of the pharmacologic management" for patients with musculoskeletal pain, and the initial choice of oral treatment for osteoarthritis[260]. NSAID are used by nearly 60% of patients for the management of musculoskeletal pain[261]. Tramadol and atypical analgesic use are usually prescribed when other analgesic medications are contra-indicated due to reasons such as co-morbidities or intolerance, or the patient has refractory pain[260,269,270]. It is likely patients would be prescribed NSAID as the first line treatment and therefore, more likely to be prescribed within the first-post operative year. Similarly, NICE guidelines recommend patients should be offered non-surgical options before any consideration of joint replacement including Physiotherapy[339]. Primary care physicians refer patients to Physiotherapy

more readily than to the Orthopaedic department and potentially reduce the number of referrals to Orthopaedics[344]. Thus, it is more likely patients would be referred to the Physiotherapy department within the first post-operative year than the Orthopaedic department. However, as with all secondary care, referral from a GP is required for patients to be seen by a Physiotherapist[333]. GP visits were unable to detect a significant difference between bearings in this study. One explanation is the significantly lower specificity of GP visits (39.6%) compared to Physiotherapy referrals (92.0%). Patients undergoing THA surgery are likely to have other co-morbidities[629] and therefore, could require significant number of visits to their primary care physician during the first post-operative year for reasons other than their surgery. It is possible the other analgesic and healthcare utilisation surrogate markers could detect differences between bearings if the measurement period was extended to two years post-operation. This would allow for those patients requiring increased management, such as Tramadol or atypical analgesics and referral to the Orthopaedic department, to be captured. The RACat cohort was unable to analyse surrogate markers measured two years post-operatively due to the short median follow up time and the significant number of revision surgeries that would be excluded.

4.5.1. Limitations

General limitations will be addressed in Chapter 5. Specific limitations of using analgesic and healthcare utilisation data are addressed in Chapter 3.

There were insufficient numbers to investigate differences between rates of Orthopaedic referral for different THA bearings. A significant difference was not detected but could exist. Physiotherapy referrals showed significant results and therefore, Orthopaedic referrals could also be able to detect differences between THA bearings as the aetiology of referrals are similar between the departments[319].

There are significant limitations regarding patients undergoing THA surgery with a MoM device within this cohort. There was no significant difference in rates of revision between MoM and other bearing devices. Therefore, this cohort may not be sufficient to investigate differences between bearing devices. The available follow-up time for the study period will likely limit analysis of possible failures potentially explaining the lack of difference in revision risk of MoM metal devices[349]. Revision rate was only able to detect MoM as an outlier after a significant period of time[204,255]. One study examining revision rates during the study period showed no differences in risk of MoM devices at that time[348], despite the expectation of there being one[628]. However, MoM devices are known to have a significantly higher rate of revision[46,59]. In addition, the crude revision rates for MoM devices were approximately double that of other bearings and the Kaplan-Meier estimates confirmed this finding. Therefore, a potential significant difference in revision rate could exist. Two surrogate markers were able to identify a difference between MoM and other bearings suggesting that analgesic and healthcare utilisation may potentially be able to identify outliers earlier than revision rate alone.

A reduced monitoring time of the RACat may limit the ability to analyse results since the study period ended at a similar time these devices were recalled[46]. Furthermore, there was significantly less uptake of MoM use in Catalan region compared to other countries. Two of the three widely used and recalled MoM devices due to increased revision rates (Bionik®, Orthodynamics, Lübeck, Germany and Icon®, Orthopaedics, Geisingen, Germany)[628], were not used in Catalonia[348]. The DePuy Articular Surface Replacement® (ASR) device, was only used in a limited number of patients within the RACat, 50 in total, prior to its recall[348]. Therefore, further investigation within cohorts with both longer follow up time and patients undergoing recalled MoM devices is required.

All other bearings that were not MoM were combined within the exposure variable. It is possible the heterogeneous control group, coupled with the different indications for revision surgery, had varying revision rates ensuring no significant difference with MoM. However, the revision rates for ceramic-on-ceramic, ceramic-on-poly, and metal-on-poly are similar and significantly different to MoM[624].

There was a significant number of patients where the bearing used was unknown with over 40% missing. Multiple imputation was not used to impute these missing values as there is limited clinical evidence the observed variables within the cohort can predict the missing observations. Subsequently a complete case analysis was performed which could represent a selection bias.

4.5.2. Strengths

General strengths will be addressed in Chapter 5. Specific strengths of using analgesic and healthcare utilisation data are addressed in Chapter 3.

A sample size was calculated for each surrogate marker. This cohort had sufficient numbers to minimise the risk of a type II error for NSAID, Tramadol, and atypical use as well as GP visits and physiotherapy referral.

This chapter represents a novel outcome measure for the investigation of differences between THA devices which has not been investigated previously.

4.5.3. Conclusion

This chapter found no significant difference in revision rate between metal-on-metal and other bearings in the RACat cohort. However, two surrogate markers were able to identify a difference suggesting that analgesic and healthcare utilisation may potentially be able to identify outliers earlier than revision rate alone.

5. Discussion

5.1. Summary of results

Chapter 2 showed there were a significantly limited number of validated surrogate markers available that adequately predicts long-term revision rate following primary THA surgery[630]. These methods included einzel-Bild-Röntgen-Analyse (eBRA) and Radiostereometric Analysis (RSA) which measure migration and wear predicting long-term revision rate within two years post-operatively. Only RSA is validated for both the femoral and acetabular components. However, RSA has significant limitations including failing to "universally predict long-term" fixation[164], and only detects aseptic loosening as the mode of failure[164,173]. Therefore, there is a need for alternative surrogate markers to be used as part of post-market surveillance. Chapter 3 investigated novel surrogate markers as part of analgesic and healthcare utilisation measured within the first post-operative year, for their association with subsequent revision in patients who have undergone primary THA surgery. The individual surrogate markers investigated included Non-Steroidal Anti-Inflammatory Drugs (NSAID), Tramadol, atypical analgesics (Tricyclic antidepressants, Gabapentin, Pregabalin), frequency of General Practitioner visits, referral to the Orthopaedic department and referral to the Physiotherapy department. All surrogates investigated were demonstrated to have significant association with subsequent revision following primary THA even when adjusting for confounders and the competing risk of mortality.

Fine & Gray (F&G) regression was used to adjust for competing risk of mortality in the final models investigating novel surrogate makers. This was decided upon a priori due to the significant proportion of elderly[483,631], and each exposure found to be significantly associated with mortality. It is recommended to report comparative results between the F&G regression and Cox proportional hazard model[632]. There were increased differences between the two regression models for the exposures NSAID use and frequency of GP visits. Smaller differences of less than five percent were found for the other exposures. In the Catalan cohort, a pre-multiple imputation association was found between all exposures and mortality. However, specificity of the exposures may explain differences between the Cox and F&G models. While commonly used for treating Osteoarthritic arthralgia, NSAID are not universally used for this purpose with other common diseases include headaches, viral illnesses such as influenza, menstrual pain and other causes of long-term pain[633]. The frequency of GP visits exposure was measured as all-cause visits due to lack of data on the reason for the consultation. There are numerous reasons a patient may consult with their GP and the measured encounters may not have represented a post-operative problem[634]. Therefore, both NSAID use, and frequency of GP visits are potentially a less specific measure of a post-operative issue in primary THA surgery. In contrast, Tramadol and atypical analgesics are used in cases of refractory and/or more severe pain, and thus with referral to the Orthopaedic and Physiotherapy departments, these exposures are more likely to be specific to the THA surgery, especially within the first post-operative year. Therefore, there would be a reduced strength of association with mortality and reduced risk of being affected by confounders that were not adjusted for in the final regression models.

Patients who had zero visits to their primary care physician within the first post-operative year were also found to be at increased risk of subsequent revision. Although a significant result was found, the p-value was close to 0.05 and may therefore be chance finding. Interestingly, a reduced proportion of those patients were either referred to the Orthopaedic or Physiotherapy departments within the first post-operative year compared to those patients with one to four visits or more than four visits. Those patients not requiring a GP visit may have direct access to secondary care without the need for referral. The British Orthopaedic Association (BOA) guidelines recommend patients be given contact information for direct access to the Orthopaedic service should any problems occur[341].

There were significant missing data in three confounding variables including Body Mass Index (BMI), alcohol status and smoking status. Multiple Imputation by Chained Equations (MICE) was used to impute missing values and create a complete dataset. Multiple imputation is regarded as the standard method to address missing data in a variety of research areas[635]. Pre- and post-MICE F&G models are reported within relevant chapters for each of the potential novel surrogate markers examined. Differences were noted for all exposures except atypical analgesics. Sub-Hazard Ratios (SHR) were between three- and 23-times larger pre-MICE compared to post-MICE and include very large 95% Confidence Intervals (CI). Atypical analgesics had a rise of SHR by approximately 20% and retained narrow 95% CI.

Discrimination was performed following the final regression model for each of the novel surrogate markers. Each surrogate marker was assessed for the extent of which

the results differ from patients who do or do not have revision[545], by plotting the receiver operating characteristic (ROC) curve[545]. A global assessment of this curve can be generated by calculating the area under the ROC curve (AUC)[545]. Unfortunately, the rating for each surrogate maker ranged from "fail" to "poor" according to commonly used gradings[546]. The AUC for the surrogate makers investigated ranged from 0.5211 to 0.6214, with referrals to the Orthopaedic or Physiotherapy department being the lowest and both NSAID and Tramadol use having the highest values. While this is above 0.5 (a reading with no discrimination), it is below the level 0.7 to 0.8 considered acceptable[636]. In addition, the Positive Predictive Value (PPV) was low (1.7%). As revision surgery is a rare outcome, the positive predictive value is expected to be low. However, when the surrogate markers were combined, the discriminatory test improved to a performance of "fair". Therefore, it is likely future use of such surrogate markers would perhaps involve devising a scoring system. One was not created here as opioids were not fully assessed but only Tramadol was included. There has been research showing opioid use in general is associated with revision risk[243]. Furthermore, the frequency of Orthopaedic visits and Physiotherapy visits has also been demonstrated to be associated with revision risk[332]. However, in the Catalan database only the binary information of whether a patient was referred was documented. Therefore, a more refined array of surrogate markers would be needed as a combined tool.

Ultimately, the methods used to investigate the surrogate markers in this thesis have attempted to reduce potential bias from confounding variables, missing variables, competing risk of mortality and censoring. Statistical significance was found for all

exposures and are unlikely to be due to chance. There was debate whether NSAID use could be causal in THA failure[606–610]. However, these failures are limited to those devices using cementless fixation technique. Following a sensitivity analysis, no significant difference was found between the cemented or cementless devices. This correlates with literature findings that NSAID use linked to reduced osseointegration is inconclusive in clinical studies[611–613] Furthermore, NSAID use is considered as an indication of "poorly functioning implant" rather than an independent risk factor[265]. There are no indications of the other surrogate markers potentially being on the causal pathway to revision.

As described earlier, there is reduced specificity with each surrogate marker and therefore, they would not directly measure post-operative issues leading to revision.

The aim of these studies was not to create a predictive model for revision but only to demonstrate a proof of concept of whether using readily available data could be used as a screening tool to identify devices at higher risk of revision. Patients identified as high risk of revision may require enhanced surveillance to investigate whether early failure occurs[319]. The surrogate markers investigated in this thesis have high negative predictive values with the lowest recording at 99.1%. Therefore, these may be used as part of a screening programme to identify potentially failing devices requiring further scrutiny using data that can easily be collected at low cost.

In chapter 4, further investigation of the novel surrogate markers was performed. The numbers needed to screen for these novel surrogate markers, measured within the first post-operative year, was estimated to range between 18 and 60. These are comparable to a widely considered gold standard surrogate marker, Radiostereometric Analysis

(RSA), in which a sample size of 15 to 25 patients per stratum is required to identify differences between devices within two years post-operation[134,136,157]. Although no significant difference between metal-on-metal and other bearings was found, two surrogate markers were able to detect a significant difference. NSAID use and referral to physiotherapy department within the first-year post-operation demonstrated metal-on-metal devices had a significantly increased risk of surrogate marker utilisation. This suggests monitoring surrogate markers could detect THA device outliers earlier and with fewer numbers compared to monitoring revision rate alone.

5.2. Clinical relevance

The findings of this thesis demonstrate a potential to create a response mechanism around the monitoring of alternative outcome measures or surrogate markers of outcome[325]. This would allow opportunities for clinicians to intervene early for those devices that are potentially failing. The novel surrogates examined here are not device specific but likely to be indicative of a post-operative issue such as pain. Furthermore, there is evidence to demonstrate increase of both analgesic and healthcare utilisation in the preceding two years of patients who undergo joint replacement[474].

There is a very clear need for novel signals of THA failure in post-market surveillance following the Metal-on-Metal (MoM) bearing disaster[46]. Significant numbers of available implants are not used in patients, potentially due to a lack of clear clinical evidence of their safety or effectiveness[117,637]. Furthermore, new arthroplasty devices introduced to the market regularly have increased risks of revision compared to

proven implants currently available[637]. One registry analysis demonstrated no evidence of improved outcome for new THA implants but nearly a third had significantly higher revision rates[637]. This study suggested an argument could be made that outcomes in these patients would have been better without the introduction of new implants. Device regulation at the time of MoM device failings was highlighted as being inadequate[46,86,638,639]. There were calls for significant change in European device regulation from both the British Medical Journal (BMJ) and the BOA[87,88]. Only a Conformité Européenne (CE) mark is required for a new device to be introduced on the European market[90]. New medical devices did not need to demonstrate clinical effectiveness to achieve a CE mark but only achieve a set of non-clinical essential requirements. In addition, a CE mark may be awarded on a claim of 'equivalency' if it is shown to be similar to an existing proven product[92]. However, even small changes in design can have deleterious effects on implant survival and patient safety as demonstrated in previous devices[94,95]. Approval from only one of the multiple bodies available[90–92] is required despite potential variation in standards. Any evidence submitted by a manufacturer to an approving body is not made public and therefore not open to scrutiny[85]. One of the public examples of the failings in MoM devices is the Articular Surface Replacement (ASR), DePuy device[46]. The ASR was introduced into the market through a claim of equivalency having not been tested in humans[46,92].

Subsequently two new regulations have come into force in 2017: the Medical Device Regulation (MDR)[186] and the In-Vitro Diagnostic Device Regulation (IVDR)[187]. Rather than redesign the current system, these new regulations serve to revise them to ensure a more modern and robust framework. The key features of the new regulation include

the requirements for clinical evidence prior to CE approval, public transparency, Unique Device Identification, and the emphasis on post-market surveillance through registries[90]. Registries can provide opportunities for long-term surveillance using national, real-world data[193], and proved to be important in providing evidence to confirm failing implants[640].

In the UK, the National Joint Registry (NJR) is one of the existing agencies that can ensure quality control, safety monitoring and allow surgeons to compare performances of devices[11]. The Orthopaedic Data Evaluation Panel (ODEP), a branch of the NHS supply chain, was formed in 2002 alongside the NJR to set a benchmark for which device performance can be compared[107]. To comply with recent changes to device regulations, a new component database has been designed in the NJR, going live in 2018, with over half of devices having a Global Trade Identification Number[641]. Registries offer an excellent resource for long-term analysis, post-market surveillance, and benchmarking[122,349]. However, both the NJR and ODEP have limitations when monitoring newly introduced THA devices to detect failing implants early. One of the key limitations is the reliance of revision as the measure of outcome[195]. A proven THA device is determined by revision rate at ten years according to current guidelines[130]. Furthermore, revisions occur after significant time has lapsed since the operation[199]. Therefore, long-term follow-up is required to ascertain the true performance of an implant. Revision surgery is not always indicative of device quality and is often not performed in patients that may require an operation, even for a failed implant, due to co-morbidities[598,642], patient choice[202], or surgeons unwillingness to recommend surgery[195]. Furthermore, devices such as resurfacing or bone conserving implants, may

be regarded as easier to revise and thus more amenable to revision surgery[195]. Revision surgery has varying types ranging from an exchange of the liner to removal of the component[202]. One of the potential safety issues regarding the use of revision as an outcome is the relatively uncommon occurrence of revision in patients undergoing THA surgery. To identify significantly higher revision rates in failing devices newly introduced in the market, large sample sizes with patients followed-up over a long-term are required[204,255]. Requiring large sample sizes results in a large number of patients undergoing invasive surgery using a failing device before being able to establish an unacceptable failure rate. This is compounded as the majority of new implants are typically introduced in small numbers initially making outliers difficult or impossible to detect[105]. Furthermore, rapid innovations and changes in surgeon's preference may make any results redundant by the time they are published[255,643].

Currently, registries analyse the revision rate of implants using Kaplan-Meier (KM) method[644,645]. The main limitation of the KM method to compare survivorship between devices is the inability to account for other explanatory variables[548].

The limitations of using revision as the primary outcome measure have been emphasised with recent MoM failings. Over one million MoM devices have been used in patients globally before their recall[205]. The ASR implant had an excellent early ODEP rating of 3A and thus deemed safe[46]. However, this device was only recalled after almost 100,000 were implanted worldwide[46,70]. Despite the mandatory national collection of data using joint registries at the time, the detrimental effects of MoM devices "were not identified sufficiently early to protect public health"[122].

Improvement of device comparative analysis involves rapid failure detection while minimising potential bias, especially from confounding by indication[122]. This requires access and analysis to databases with detailed information. While prospective data collection through registries would offer the most robust dataset, there is the potential for reduced accuracy or completeness for information outside the immediate scope of an operation or the device[122]. Joint registries globally provide details related to both operation and device, but lack crucial information on patient characteristics and outcomes other than revision[122]. There are significant factors such as obesity[122] or smoking[646] associated with post-operative adverse events. Registries are likely to form an integral part of national healthcare systems and need to be improved as they are relied upon to provide evidence for policy decision making in removing a device from the market[46]. There has been expansion of data collected in other registries such as complications, readmissions, early revision rates[647,648], but crucial data on potential confounders of revision are not collected. Furthermore, the NJR still does not collect data that other registries do such as dislocation or peri-prosthetic fracture, medical complications, drug use post-operatively (analgesic and antibiotic), and healthcare utilisation such as length of stay or discharge destination[231]. There is data on age, sex, BMI, marital status, deprivation index, diagnosis, previous surgery, and Charlson score. However, it is missing socioeconomic status, lifestyle factors (smoking, level of activity, alcohol abuse), medication use and other co-morbidities that have been shown to be associated with revision[231].

It is possible that with adequate registry data and analytical methods surgeons would cease using failing devices early without the need to wait for legislation[649]. However, despite its potential, registry data is under used due to barriers such as limited access

from legal restrictions, insufficient numbers of staff trained for research, and in some cases lack of awareness of its existence[650].

There is a need to adopt strategies similar to drug surveillance that have the potential to improve on post-market safety surveillance[651,652]. There is recognition for the need of "a more rigorous system" due to the limitations of current post-market surveillance practices[319]. There have been calls to evaluate medical devices using real-world data that is already routinely collected[319,653]. There has been a surge in the number of available THA devices in the market from 62 in 1996 to now over 2,800 different combinations currently available[102]. This rise in the number of new hip implants do not seem to address any new concerns or ensure improved outcomes[637]. A report conducted for the BOA highlights the need for 'implants that demonstrate survival rates of at least 90% at ten years' as the 'gold standard' and argues that 'offering patients more expensive implants with little or no added benefit denies other patients Orthopaedic care'[600]. However, after recent analysis this benchmark has moved to consider a gold standard implant having a <5% revision rate at ten years[130]. Priority should be given to monitoring devices in real-time by investigating specific risks[654]. New outcome measures are needed that offer improved measures of benefits versus risk assessments and cost-effectiveness[655,656], including pain, disability, and satisfaction[122]. Pain, for example, has been identified as an early indicator of post-operative problems increasing the risk of subsequent revision[657–659].

Therefore, to obtain broader data crucial to expanding device comparative analysis is heavily relied upon linkage to other datasets[122]. Such linkage allows researchers investigating new devices to analyse early signals of THA failure, adjust for known confounders and potentially determine causation[122]. A recent example being the

linkage of the UK primary care database and the NJR to analyse cancer risk in patients with MoM devices[660].

Linkage offers the potential to address fundamental flaws in the current surveillance system of THA devices due to the inability to monitor adverse events in 'real-time', the lack of comparative studies to draw conclusions, and minimal data collected for patient characteristics and several others[122,194]. Real World Evidence (RWE) is generally described as evidence derived from studies examining data generated or existing outside formal clinical trials such as randomised controlled trials (RCT)[661]. It is increasingly used to provide greater validity and robustness in the assessment of new medical technologies[591]. Furthermore, the use of RWE is becoming increasingly prevalent in appraisal of medical devices submitted to NICE[591]. RWE can determine whether findings from an RCT are in keeping with real clinical practice conducted by varying institutions, clinicians and population groups[662]. When RWE forms part of the evidence base for CE mark approval it is more positively received by appraisal committees[591]. It is recommended by the European Medicines Agency as part of a complete assessment for new devices[591]. Electronic medical records (EMR) incorporate patient level data collected systematically and digitally stored. Such data can be used to improve patient care or outcome with aspects made available for research as anonymised datasets[591]. Advances in technology have increased the range and depth of EMR data that is routinely collected, making it possible to investigate potential surrogate markers of THA failure or alternative endpoints to revision for post-market surveillance. In addition to monitoring pain, increased frequency of healthcare utilisation has been identified as a potential alternative marker of THA outcome[332].

There is an increasing trend to use EMR data in observational studies as "a source of real-world evidence" in assessing safety and patient outcomes for medical devices[659,663]. Their use in post-market surveillance of pharmaceuticals have shown to improve accuracy of signal detection[664]. EMR data have significant benefits including continuous data capture and wide coverage of patient population (national in countries such as the UK[321,322]) over a prolonged follow-up period[659]. Therefore, EMR data offers a potentially valuable resource for real-world medical device surveillance which could complement existing structures such as the NJR. The use of such indicators monitored using EMR data can provide a novel, cost-effective method for post-market surveillance of THA devices in large cohorts of patients improving safety. This could be extended to an almost automated system further reducing surveillance costs with both analgesia and healthcare utilisation providing the framework through EMR databases. Subsequently, this could provide the framework for national medical device surveillance.

In an increasingly digitalised setting, pharmaceutical records and joint surgery data are routinely recorded and stored. In an integrated healthcare system, these records can be linked to patients who have had THA surgery to enable post-market surveillance of implants[258]. The studies in this thesis have demonstrated the potential for several surrogate markers that are routinely collected and can be analysed through linkage of primary care and joint registry databases.

5.3. Generalisability

5.3.1. Joint registry

Despite a regulation introduced in 2003 by the Spanish Health Ministry for the creation of a National Implant Registry, one does not exist that covers the National Health System[348]. Currently, there are several registries formed through initiatives from individual scientific societies across varying regions in Spain[348]. One of the most comprehensive and internationally recognised regional joint registry in Spain is the Catalan Arthroplasty Registry (RACat)[231,665]. The RACat is one of only a handful of European joint registries that are regional based[231], but may potentially be representative of Spain and therefore generalisable to a national registry. A study investigating national hospital data during a similar period to the studies here demonstrated 45% of men and 55% of women had undergone THA surgery[553], compared to 52.9% of women in the RACat[231]. Similar proportions of patients in the RACat suffered from Osteoarthritis as the indication for surgery with 77.9%[231] compared to 75% in Spain as a whole[553].

A potentially significant difference between the Catalan registry and Spanish data was the distribution of patients according to their Charlson score. In Spain, a far greater proportion of patients (78.5%) undergoing primary THA had a score of zero, 18.4% with a score of one to two and only 1.1% scoring over two[553]. In contrast, the RACat reports only 44.6% of patients had a score of zero. This suggests that the cohort of patients in the study of this thesis have significantly greater co-morbidities that could potentially bias any results when generalising to the general population. However, a sensitivity analysis excluding Charlson score as a confounding variable was

performed. The associations found in the main analysis between exposures and subsequent revision remained almost identical.

The National Joint Registry (NJR) of England and Wales is a well-established joint registry and one of the largest in the world for hip replacement, providing an excellent source of data[231]. National arthroplasty registries are restricted to those countries covering a population of less than 20 million with the exception of the UK and Germany[231]. This makes the NJR a good comparator for the RACat with both registries established in 2002 and 2005 respectively[231]. The NJR is over 17 times larger in number of patient participants than the RACat for hip replacement[231]. Since its inception initial compliance to submitting data by manufacturers and institutions was poor in 2004 at approximately 40% and rose to 80% in 2006[102]. Now there is almost universal coverage of all primary and revision THA operations being performed within the United Kingdom following mandatory requirement for submission of data[104]. Private sector compliance was mandatory since the formation of the NJR[102]. Coverage of THA procedures is 97% for all THA operations[231], with some reports placing the NJR coverage as high as 99%[348]. In comparison, RACat coverage for Catalonia is between 80% to 85%[231]. Furthermore, RACat coverage was marginally over 70% during the study period[348] compared to over 99% in the same period for the NJR[666]. Therefore, the NJR can be considered highly representative of all types of patients, health professionals and institutions[556].

The NJR and RACat have similar mean ages of 68.5 and 68.8 years respectively for patients registered having primary THA surgery[231]. In addition, the percentage of

women undergoing hip replacement is similar between the two registries with the NJR reporting 59.8% and the RACat 52.9%[231]. The mean Body Mass Index (BMI) for NJR is 28.7[231] in comparison to the RACat with an almost identical mean of 28.2 during the study period.

There are differences in the fixation type between both registries. The NJR reports of the primary THA operations performed 34% were cemented, 39% cementless and 20% hybrid fixations[556]. In contrast, during the study period, the RACat demonstrated 21.4% were cemented, 47.6% cementless and 14.5% hybrid fixations. This suggests a potentially significant difference in device selection between the two countries affecting generalisability.

Both registries also have a comparable crude risk of revision for the reporting period of the studies conducted in Chapter 4. The risk of revision at three years for THA was 2.9% in the RACat, and 2.32% in the NJR[348].There are several factors that could have caused the higher revision rate seen in the RACat, such as implant selection or poor quality of data in the early phase following the introduction of the registry[349]. However, these figures are crude without adjusting for potential confounders and thus caution is necessary when comparing risk of revision from raw data[349]. Furthermore, there is a potentially significant difference between the number of patients registered with primary Osteoarthritis in the NJR documenting 92% compared to the RACat with 78%[231]. This difference could potentially bias any future studies investigating new devices whose indication will likely be for Osteoarthritis. Another difference between the NJR and the RACat is significantly more patients undergo THA for hip fracture in Catalan compared to the UK. In the NJR the vast majority of patients are treated with hemiarthroplasty with only 3% of primary THA operations performed for fractured

neck of femur[556]. In contrast, nearly a third of patients in the RACat cohort had primary THA surgery to treat a fracture during the study period. To ensure that indication would not confound the effect size, a sensitivity analysis was undertaken to include only patients undergoing THA for indication of Osteoarthritis. The associations found in the main analysis between exposures and subsequent revision were stronger.

The UK has a higher rate of operations and are therefore more likely to be experienced surgeons. When comparing to population numbers the UK has an annual rate of primary THA operations of approximately 0.14%[556,667] compared to the Catalonia data with a rate of approximately 0.06%[349]. Therefore, the NJR has a rate of primary THA operation double that of the RACat. In the evaluation of new implants, the learning curve could impact revision rates for patients operated on during this period. Therefore, in the early evaluation of implants surgeon experience could be an important factor that influences bias.

Both registries collect patient, device, hospital, and surgery level data with the NJR having one of the "highest granularity of outcome data and covariates"[231]. Patient related factors include age, gender, BMI, marital status, deprivation index, diagnosis, previous surgery, and Charlson score for the NJR. The RACat lacks marital status, previous surgery and Charlson score but has Elixhauser score and medication use. The NJR collects both hospital volume and type whereas the RACat collects hospital volume data only. The RACat collects no surgery related factors such as surgical approach, operation time, surgeon experience, anaesthesia type, antibiotics, anticoagulation which are all collected by the NJR. Both registries collect device related factors including implant details, fixation type, and bearing type. Several arthroplasty

registries in addition to the RACat, including Denmark and Sweden, have moved to incorporate data collection of surrogate measures of complications such as drug prescription[231].

There are differences in the outcomes reported by either the NJR or the RACat[231]. As with almost all European arthroplasty registries, revision rate is the primary outcome measure of device survival. In addition, the NJR reports infection, mortality, Patient Reported Outcome Measures (PROM), pain, function, general health through assessments, satisfaction, and healthcare utilisation only including readmission and healthcare costs. In contrast the RACat collects mortality, pain, function, and post-operative analgesic use only in addition to revision rate.

Despite the considerable collection of co-variate and outcome data there are significant limitations with using registry data alone to investigate new potential surrogate markers of revision rate in primary THA using observational studies. In the current format of European registries, poorly performing devices are detected through revision rate only, despite other outcome measures being collected. The systematic review in Chapter 2 revealed numerous studies identifying a number of factors associated with revision rate not routinely collected by registries[120]. Observational data such as the NJR or RACat alone is limited by confounding variables. For example, surgeons may choose specific type of prosthesis or surgical approaches based on a patient's characteristics or simply on the surgeon's expertise. It is important to adjust for confounding factors when conducting observational studies and thus registries need to merge their data with other sources[122]. Unfortunately, a study investigating the association of cement and mortality in primary THA concluded the NJR does not

contain adequate details of potential confounders such as comorbidities, socio-economic status, lifestyle factors (including smoking and alcohol status), and pharmaceutical use, all of which have known impact on THA device longevity[668]. In order to use pharmaceutical use as an indicator of failure, high linkage between a joint registry, primary care and hospital databases is required. The NJR is poorly linked to primary care databases that contain data on potential confounding factors such as the Clinical Practice Research Datalink (CPRD) in the UK. Following a government directive one study used a collaboration between the NJR and CPRD to investigate the risk of cancer and MoM devices which was only able to obtain 40% linkage between datasets[660]. Therefore, poor linkage with primary care data could result in registries lacking sufficient variables to adequately build an analytical model. In addition, revision rate is a rare outcome requiring larger numbers of participants in comparison groups compared to what may be required in a clinical randomised controlled trial. Thus, registries with lower linkage may not have sufficient power to investigate any effect size and may introduce bias if such a low proportion of the population is included within the cohort.

In contrast, the RACat has excellent linkage and is one of the only arthroplasty registers to have a database which is highly linked to national primary care records, pharmacy invoice data, mortality registry and the hospital episodes data. Linkage is more than double the NJR at approximately 85% allowing for the adjustment of potential confounders[346]. This allows for a detailed analysis to investigate pharmaceutical and healthcare utilisation as potential indicators of THA failure[348]. Furthermore, the RACat is able to analyse a cohort of over 17,000 patients with linked records. Other studies investigating the association between analgesic drug use and subsequent revision in primary THA have had similar cohorts but with fewer patient

numbers compared to the RACat. Namba et al investigated the association of opioid use and THA revision in the Kaiser Permanente cohort with a total of 12,859 patients evaluated between 2008 and 2011[258]. In addition, healthcare utilisation (primary care visits and orthopaedic visits) was investigated in the same cohort[332].

The only recorded linked cohort between the NJR and primary care data was undertaken by Lalmohamed et al in 2013 investigating potential cancer risk in MoM hip replacements[660]. This UK linked cohort had over 11,500 patients compared to the 17,000 of the RACat during a similar study period.

5.3.2. Healthcare system

There are similarities between the UK and Spanish healthcare systems. The Spanish healthcare system, like the UK, is a National Healthcare Service financed through governmental taxes and both healthcare encounters and hospital admission being provided free of charge[349]. In a similar manner to the UK, while all aspects of care is provided for free, patients are charged for pharmaceuticals[378]. Furthermore, Spain has a comparable healthcare system to the UK with primary care acting as the gate keepers for secondary care[350,669].

However, the Spanish healthcare system is primarily decentralised consisting of 17 regions with complete responsibility for financial decisions and service delivery[378]. Although a national health technology assessment body exists in Spain, economic evaluation of new technologies is not mandatory. In addition, there is no national body explicitly conducting cost effective analysis to aid decision-making process of what is provided in the healthcare service[378]. In contrast, the UK has only three national bodies

evaluating cost-effectiveness using the same set of thresholds, with the National Institute for Health and Care Excellence (NICE) having its own framework, a wide scope including both pharmaceuticals and technologies and their process is not limited to submissions by manufacturers[378]. This could allow for newer devices to be used more freely within the national health service in Spain compared to the UK. However, NICE has a track record of influencing prescribing patterns rather than the medical device use or surgical procedures[670].

The Spanish healthcare system is part of a national health service and access to secondary care is only through primary care[350]. The studies conducted investigating potential surrogate markers in this thesis are likely generalisable to healthcare systems which are nationally funded with primary care physicians acting as the gate keepers. However, the results of these studies may not apply to those with a predominantly private system, or those that allow direct access to secondary care. There can be differences between populations of various countries and their structure available in the database. There are differences in coding within databases including healthcare practices and guidelines in addition to methods of financing. All of these could ensure generalisability is not possible and there may be a need for bespoke outcome measures in each country. For example, time to joint replacement can have multiple factors including surgeon preferences, disease severity resulting in the UK having longer median time to surgery compared to the United States[671]. The World Health Organisation has suggested the variation in rates of hip replacement across the European Union can be mostly accounted for in the differences in healthcare reimbursement practices[672].

5.3.3. Primary care database

The UK primary care database, Clinical Practice Research Datalink (CPRD), is considered one of the most comprehensive in the world[591]. There is virtually universal coverage of all primary care practices in the UK since 1996[321,322]. The CPRD collects data similar to the Spanish primary care database (SIDIAP)[346,352] including patient demographics, encounters, diagnoses, symptoms, prescriptions, investigations among others[590]. At approximately the same time period of the SIDIAP cohort used in this thesis, the CPRD held over 11 million patient records[590]. In comparison, the SIDIAP holds records of over 5.5 million, approximately 74% of the Catalan population[352]. This is significantly higher compared to the coverage of CPRD which was approximately 7% in 2015[590]. Furthermore, data linked to other databases such as hospital episode statistics and office of national statistics is lower for the CPRD with less than 60% linkage. SIDIAP is highly linked, approximately 85%[346], to other databases.

5.4. Limitations

5.4.1. Limitations of database

Quality of data is a common issue with analysis of primary care databases and with the majority of studies analysing any real world database[591]. During the initial phase following its introduction, reporting for RACat was on a voluntary basis leading to poor coverage and data quality. The RACat coverage is considerably lower than other joint registries during the study period from as low as 24% in 2005[348], which could lead to bias due to systematic differences between those hospitals reporting or not reporting to the registry. However, data quality improved during the initial stages of data

collection and thus the study period has better coverage and data quality in the later part of follow-up[231,348,349].

Databases such as the joint registries and primary care databases were "not constructed for research purposes"[671], but rather to capture real-world experiences at a patient level. Therefore, details potentially pertinent to research purposes, such as severity of disease in osteoarthritis, functional disability, reasons for encounter or referral are not always coded. The use of electronic medical records data in research may be limited when investigating risk factors in medical devices[671].

The data used is administrative and therefore, is subject to potential errors from coding or under reporting that could lead to bias. It is possible data was misclassified or encounters are missing due to poor coding. It is also possible there is a lack of laterality to the data used resulting in revision surgeries incorrectly linked to primary surgeries[673]. However, there are quality checks to ensure high laterality[348,349]. In addition, it is unlikely that these factors would cause systematic differences between the groups analysed and therefore, likely to cause minimal effect on estimations.

Another limitation is the lack of validation for each recorded event. Information is gathered using standardised ICD-10 codes and structured spreadsheets. Only practices that achieve quality control standards can contribute to the database[354]. This may lead to bias with the exclusion of practices with poor quality control. However, the RACat and SIDIAP data has previously been used in several epidemiological studies and a number of outcomes have been successfully validated in data extracts from this source[355–360].

Large data registries can suffer from missing data that could lead to bias if excluded. The cumulative effect of missing data in several variables could lead to the exclusion of

a substantial proportion of the original sample. This could lead to a significant loss of power.

5.4.2. Limitations of main analysis

Data from the Catalan linked database is strictly observational. Therefore, only associations between exposures and subsequent revision rate were reported as no inference of causality can be made. There are other known potential confounders which were not measured and thus not accounted for within the regression models such as length of operation or length of hospital stay[553,674].

There was a lack of detail with the exposures that can cause challenges when interpreting results of possible associations. These include the reasons for GP visits, or referrals to the Orthopaedic and Physiotherapy departments. It is possible patients were visiting their primary care physician or referred to secondary care for reasons other than their primary THA surgery. However, the aim of these studies was to identify possible early signals of failure that are routinely measured and thus providing a cost-effective screening tool.

There was a lack of information regarding patients who were offered revision surgery but subsequently declined or were deemed unfit due co-morbidities. These patients could have systemic differences from those who have undergone revision surgery leading to bias. Willingness to undergo THA surgery is the strongest predictor of subsequent operation which is challenging to measure[335]. However, multiple risk factors of poor outcome have been identified that can influence the patient or surgeon's decision to continue with a hip replacement surgery. These were included as

confounders in the main analysis. Furthermore, the use of revision rate alone as a measure of device survival will not identify these patients but analgesic and healthcare utilisation may be sensitive, and therefore potentially provide more validated findings of outcome in THA devices.

The studies were unable to fully analyse the reasons for revision due to missing data. Although revision commonly occurs due to aseptic loosening and therefore is likely related to product design, there are other causes of revision[365]. The majority of revisions in this cohort were indicated for aseptic loosening. However, some were for periprosthetic fracture or infection which have other risk factors that may not be device related[131]. These can have an impact on implant survival and include surgical technique, surgeon, hospital, patient factors, and country. As a result, this further highlights revision rate as a crude endpoint in the assessment of performance for implants.

There is no information regarding nationally recommended or common patient rehabilitation protocols for patients undergoing THA surgery in Spain. Recommendations from the BOA and NICE were used to establish potential surrogate markers. Surgical practices can vary between hospitals and surgeons which could be a potential source of residual confounding. However, NICE guidance is highly regarded in Europe[378] and may be adopted by a number of institutions that lack national guidelines. Furthermore, the RACat and SIDIAP databases has good coverage of the entire Catalan region and incorporates a mixture of hospitals and surgeons with differing experiences and practices[348].

As with all observational studies, the Fine and Gray regression modelling relies on the identification and collection of data for potential confounders. It is possible that a systematic difference is present even after adjusting for multiple confounders resulting in residual confounding[547]. Identifying potential confounders may be more challenging when there is a high correlation between covariates[547]. Sensitivity analysis was performed on indication, fixation type and bearing. However, there were no specific interactions identified a priori to be investigated.

The positive and negative predictive values are calculated and thus dependant on the prevalence of the disease[544]. In the case of revision, with a very low prevalence, the positive predictive value will be poor even if the sensitivity or specificity are high. This is due to the positive predictive value not being close to one when the prevalence is low. The proportion of patients requiring revision who were in the high-risk categories for each surrogate marker investigated was very low ranging from 1.3% to 3.3%. This results in the inevitable false positive results of any test used[544]. After calculating the area under the curve, maximising the sensitivity and specificity may not necessarily be the best approach with other factors such as cost, implications of false results and the prevalence of the disease. In the case here, the exposures investigated are not individually a diagnostic tool and thus the area under the curve is a broad representation of its discriminatory power. The prevalence of revision is low and for the exposure to adequately detect these patients would be at the cost of a significant number of false positives[545].

5.4.3. Limitations of multiple imputation

In general, real-world data is not collected for research purposes, it can suffer from missing data and bias as a result of primary care being episodic and reactive in nature. Three confounding variables had significant proportions of missing data despite regular monitoring of data quality and completeness being required to limit missing data[635]. Missing data could lead to substantial bias due to complete case analysis[675]. Multiple imputation was used to address the missing data. However, not all the variables required to impute missing data are necessarily measured. There could be potential variables that are unmeasured with better predictive power of missingness that could change the distribution of the confounders resulting in bias.

Furthermore, the risk of bias from missing data may depend on their reasons as to why the data is missing. Despite the assumption missing data is categorised as Missing At Random (MAR) and not Missing Completely At Random (MCAR) or Missing Not At Random (MNAR), this remains only an assumption[635]. It is not possible to definitively ascertain the type of missingness from the observed data in regard to whether they are MAR or MNAR[676]. However, it is possible for the observed data to give some indication whether these values are MCAR[677]. In this study, the characteristics of patients with missing values were tabulated and evaluation showed that these variables meet conditions of MAR.

Unfortunately, there will be some data that inherently retain systematic differences despite accounting for other variables as it is impossible to account for all the systematic differences using observed data alone[675]. Using prior knowledge and clinical expertise it is likely that the variables used for imputation are sufficient. Comparisons were made between complete case analysis and analysis following multiple

imputation. While differences existed between the results, there were no opposing results. Furthermore, there are no clinical reasons to suspect systemic differences in reporting of these confounders such as Body Mass Index (BMI)[484]. Although most revisions were recorded with an indication as aseptic loosening, approximately 20% were for peri-prosthetic infection. BMI is known to increase the risk of post-operative infection and reduced reporting may be distributed differently in BMI classes. This could lead to a higher risk of revision being associated with increasing BMI being underestimated[484].

Finally, each confounding variable with missing data had potentially significant proportions of patients who had revision surgery which could lead to systematic differences. However, the multiple imputation technique should aid to minimise this bias.

5.5. Strengths

On balance, the methodology and the data used ensured there is likely a low risk of bias during the investigation for associations between potential surrogate markers and subsequent revision rate. There is inclusion of a significant number of unselected patients undergoing THA with good coverage of the sample region and high completeness[351]. The quality of clinical information of patients is continually monitored and only general practitioners who achieve quality control standards can contribute to the database[354], and therefore the database is considered good quality. The clinical information is rich, prospectively collected and comparable to other joint registries in Europe[231]. This is coupled with high linkage to primary care records using a unique

patient identifier leading to high accuracy. The linked cohort between registry and primary care databases is larger in comparison to a linked UK cohort of the same time period[660]. There is generalisability of patients, surgeons and hospitals included in the cohort. The data was prospectively collected and predefined with structured templates for data collection, leading to potentially high internal validity of the data.

5.6. Future work

One of the significant problems with big datasets is both availability and analysis in the discovery of early signals for safety monitoring[678]. Furthermore, understanding the potential markers and focusing them into actionable information is challenging[678]. It is imperative "the right data is delivered to the right people at the right time"[678] to ensure safety when introducing new devices. The studies of this thesis are primarily exploratory. Further studies are required to conduct multiple hypothesis adjustments[243], determine cut off for medication use and establish validity of findings. The identification of surrogate markers was limited to the availability of data. Other studies have investigated opioid use in general rather than Tramadol use alone[243,326]. Therefore, additional studies are required to identify other potential surrogate markers and whether the combination of these can be used to increase diagnostic power.

Pain is perhaps the underlying biological process the surrogate markers are measuring, and it has been suggested it is a more appropriate outcome measure of THA surgery[202]. An association with patient reported outcome and subsequent revision rate in primary THA has been demonstrated[221]. Although there is potential to combine PROMs with other surrogate markers, their use as an outcome measure will need further

investigation. Particular focus on a "minimum clinically important difference"[679] is required to establish effect size in future post-market screening studies.

Further studies are also required to determine whether the surrogate markers identified can be used as outcome measures during the surveillance of new THA devices to identify those with a higher revision rate early. This thesis demonstrated two surrogate markers had the potential to identify failing bearings earlier than revision rate alone. However, the analysis potentially lacked power and an appropriate number of patients undergoing known failing MoM devices. Any potential surrogate marker will need to be validated against the identification of known failed devices but would require a highly linked cohort, such as the RACat, to adjust for potential confounders. These would need to be comparative studies against proven devices.

The importance of post-market surveillance in assessing device safety using real-world evidence has been highlighted and remains a challenge for stakeholders such as manufacturers, regulatory bodies and healthcare providers[659,680]. Despite the UK having a national primary care database, one significant issue is the fragmentation of structured records between secondary, joint registry and primary care. If the use of routinely collected real-world data can be used for post-market surveillance of medical devices, methods will need to be formulated to ensure analysis of this data is efficient and in real-time. The recent failings of MoM devices highlight the potential need for automated methods of post-market surveillance[681]. One potential solution is the use of machine learning[659] to identify important outcomes such as pain[221]. Callahan et al demonstrated, using machine learning in EMR data, an association between hip pain in the first year post-operatively and risk of subsequent revision[659]. Using such methods

would enable an automated surveillance in "near real-time" and "augment existing registries"[659].

EMR systems have large sample sizes, including national coverage in a number of countries, and long observational periods. Therefore, it is highly probable that EMR data will be utilised in studies assessing safety and comparing effectiveness for both pharmaceuticals and medical devices[671]. Granularity of data is lacking in EMR systems and their use in research may be limited when investigating risk factors in medical devices[671]. Future research may need to incorporate the development of coding algorithms suitable in joint replacement population that can explore clinical factors which in turn can reliably be used in observational studies from EMR data[671]. However, the pressures of cost in follow-up of surgical patients including conducting clinical trials, makes the use of EMR data attractive. It is anticipated that EMR databases "will be increasingly used for epidemiologic purposes"[671].

Current methods of detection are empirical in nature and combined with clinical judgement to identify failing devices with increased revision risks[682]. However, this method primarily fails to address the issue of confounding due to patient risk factors. Methods to address both censoring inherent to survival analysis and confounding from patient characteristics are sought[640].

One emerging methodological technique showing promise for post-market surveillance of THA devices using revision rate as an outcome measure is the use of machine learning[640]. However, this study has demonstrated machine learning is unable

to universally detect known failing implants using revision as an outcome. In addition, these devices were undetected due to the limited number of revisions within the cohort, and thus would potentially perform poorly in the context of a newly introduced implant with a small cohort initially. Therefore, alternative outcome measures or surrogate markers of revision are required such as those explored in this thesis. Further studies should be conducted to investigate whether machine learning can detect known failed devices using alternative outcome measures.

Another methodological technique suggested is the use of propensity scoring to reduce the impact of confounders[683]. This method builds a score based on the conditional probability of an individual having the exposure given a set of known covariates[684]. Randomised controlled trials are considered the highest level of evidence, but are expensive to conduct, and may still be biased causing a lack of evidence from such trials[683]. Therefore, clinicians commonly rely on observational studies. The use of propensity scoring methods are considered more appropriate for comparative studies that aim to confirm safety and effectiveness[685,686]. However, they have been shown to perform poorer than conventional regression modelling such as Cox proportional hazard or Fine & Grey analysis, in larger databases[687]. Furthermore, given the low numbers of patients undergoing THA surgery with a newly introduced implant, and the rarity of revision, alternative outcomes to revision may still be required to adequately identify failing devices early.

Future studies will rely on adequate linkage of datasets. Combining data is required to address the initial low cohort of patients undergoing THA surgery with newly introduced devices. There is increasing consensus to adequately detect failing devices

early and the acknowledgement that large, multinational registries through linkage are essential[122,688,689]. Novel surrogate measures that are more widely used, such as drug and healthcare utilisation, will help alleviate the rarity of revision initially that hinders the early identification of any significant differences.

Despite larger cohorts through linkage of international registries, any analysis will continue to be limited from potential confounding variables. Even with machine learning, adjustment from known confounders and multiple imputation for missing data is required[640]. Therefore, linkage between primary care records and joint registries are required at both a national and international level.

Whichever method or outcome measure is used, it is likely any solution to the issue of safety for new devices will rely on a combined approach of pre-market clinical evidence, post-market surveillance used as a screening tool to identify potentially failing implants, and the comparative studies investigating survival and effectiveness.

5.7. Conclusion

There is a rapidly growing number of devices available with new introductions having significant variations[690], aimed primarily at reducing complications, costs and improving patient statisfaction[691]. However, these implants often have no evidence demonstrating any improvement compared to proven devices, and furthermore, approximately a third cause significantly worse outcomes[637]. Nieuwenhuijse et al concluded there is insufficient "high quality evidence" of improved benefit compared to proven devices and alternatives that are known to be safe[692]. They also stated

existing, well proven, implants may be safer. MoM devices have not only been shown to cause harm but have demonstrated no benefits when compared to metal-on-polyethylene bearings[693]. While there needs to be a balance between delay to market for a device and the avoidance of harm from those performing poorly there is a need for improved requirements of data on safety and effectiveness[693].

Modern technology and the widespread use of EMR systems has led to the opportunity to potentially use such rich data as a proxy for THA device outcome. This is very much a reality in the UK where there is almost universal EMR coverage nationally. The simple assessment of analgesic and healthcare utilisation alone can provide valuable data regarding the post-operative period and identify early deviations from the normal rehabilitation in potentially failing implants. Using large cohorts of patients, this type of surveillance could assist stake holders with monitoring patient outcomes using these basic metrics.

It is widely considered that large longitudinal databases hold epidemiological data with the highest quality available at a patient level and it is unlikely other sources will emerge[591]. Joint registries lack information on potential confounding variables[122] and primary care datasets lack secondary or tertiary care information[591]. Furthermore, primary care databases are unlikely to capture significant data regarding new devices, especially those used in secondary care[591]. Therefore, linkage of datasets is vital as no singular database contains all variables required for the comparisons of devices using any validated surrogate marker.

Despite the potential of surrogate markers, those investigated here are demonstrated to be associated with only short to mid-term revision rate, it is likely that these findings would remain useful and potentially valid as a "safety indicator for later failures" in primary THA surgery[319]. The surrogate markers of THA outcome investigated in this thesis can potentially be used to identify early problems for new devices and determine high-risk devices that require more active surveillance[319].

6. Bibliography

1 Malak T, Beard D, Glyn-Jones S. Total Hip Arthroplasty: Recent Advances and Controversies. *Arthritis Res UK Top Rev* 2014; **Spring 201**.

2 Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet. 2007. DOI:10.1016/S0140-6736(07)60457-7.

3 Charnley J. The Bonding of Prostheses to Bone by Cement. *J Bone Jt Surg Br* 1964; **46**: 518–29.

4 Liu XW, Zi Y, Xiang LB, Wang Y. Total hip arthroplasty: A review of advances, advantages and limitations. Int. J. Clin. Exp. Med. 2015.

5 Charnley J. The long-term results of low-friction a primary arthroplasty intervention. *J Bone Jt Surg [Br]* 1972.

6 Wilcock GK. Benefits of total hip replacement to older patients and the community. *Br Med J* 1978. DOI:10.1136/bmj.2.6129.37.

7 O'Boyle CA, McGee H, Hickey A, O'Malley K, Joyce CRB. Individual quality of life in patients undergoing hip replacement. *Lancet* 1992. DOI:10.1016/0140-6736(92)90673-Q.

8 Laupacis A, Bourne R, Rorabeck C, *et al.* The effect of elective total hip replacement on health-related quality of life. *J Bone Jt Surg - Ser A* 1993. DOI:10.2106/00004623-199311000-00006.

9 Fordham R, Skinner J, Wang X, Nolan J. The economic benefit of hip

replacement: A 5-year follow-up of costs and outcomes in the Exeter Primary Outcomes Study. *BMJ Open* 2012. DOI:10.1136/bmjopen-2011-000752.

10 Bjorklund L. The Bone and Joint Decade 2000-2010. Inaugural meeting 17 and 18 April 1998, Lund, Sweden. *Acta Orthop Scand Suppl* 1998.

11 National Joint Registry 15th Annual Report. 2018.

12 Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK. Future projections of total hip and knee arthroplasty in the UK: Results from the UK Clinical Practice Research Datalink. *Osteoarthr Cartil* 2015. DOI:10.1016/j.joca.2014.12.022.

13 Cassidy RS, O hEireamhoin S, Beverland DE. Guidelines for the follow-up of total hip arthroplasty: do they need to be revised?. *Bone Joint J* 2019. DOI:10.1302/0301-620X.101B5.BJJ-2018-0853.R2.

14 Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the u.s., 2014 to 2030. *J Bone Jt Surg - Am Vol* 2018. DOI:10.2106/JBJS.17.01617.

15 Patel A, Pavlou G, Mújica-Mota RE, Toms AD. The epidemiology of revision total knee and hip arthroplasty in England and Wales: A comparative analysis with projections for the United States. a study using the national joint registry dataset. *Bone Jt J* 2015. DOI:10.1302/0301-620X.97B8.35170.

16 Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Jt Surg - Ser A* 2007. DOI:10.2106/JBJS.F.00222.

17 Ackerman IN, Bohensky MA, Zomer E, *et al*. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. *BMC Musculoskelet Disord* 2019. DOI:10.1186/s12891-019-2411-9.

18 Murphy LB, Helmick CG, Schwartz TA, *et al*. One in four people may develop symptomatic hip osteoarthritis in his or her lifetime. *Osteoarthr Cartil* 2010; **18**: 1372–9.

19 Metcalfe D, Peterson N, Wilkinson JM, Perry DC. Temporal trends and survivorship of total hip arthroplasty in very young patients. *Bone Joint J* 2018. DOI:10.1302/0301-620x.100b10.bjj-2017-1441.r2.

20 Schmitz MW, Busch VJ, Gardeniers JW, Hendriks JC, Veth RP, Schreurs BW. Long-term results of cemented total hip arthroplasty in patients younger than 30 years and the outcome of subsequent revisions. *BMC Musculoskelet Disord* 2013. DOI:10.1186/1471-2474-14-37.

21 Mäkelä K, Eskelinen A, Pulkkinen P, Paavolainen P, Remes V. Cemented total hip replacement for primary osteoarthritis in patients aged 55 years or older: Results of the 12 most common cemented implants followed for 25 years in the finnish arthroplasty register. *J Bone Jt Surg - Ser B* 2008. DOI:10.1302/0301-620X.90B12.21151.

22 Berry DJ, Harmsen WS, Cabanela ME, Morrey BF. Twenty-five-Year Survivorship of Two Thousand Consecutive Primary Charnley Total Hip Replacements. *J Bone Jt Surgery-American Vol* 2002. DOI:10.2106/00004623-200202000-00002.

23 Evans JT, Evans JP, Walker RW, Blom AW, Whitehouse MR, Sayers A. How long

does a hip replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet (London, England)* 2019. DOI:10.1016/S0140-6736(18)31665-9.

24 WILES P. The surgery of the osteoarthritic hip. *Br J Surg* 1958. DOI:10.1002/bjs.18004519315.

25 Charnley J. ARTHROPLASTY OF THE HIP. A New Operation. *Lancet* 1961. DOI:10.1016/S0140-6736(61)92063-3.

26 CHARNLEY J. Anchorage of the femoral head prosthesis to the shaft of the femur. *J Bone Joint Surg Br* 1960. DOI:10.1302/0301-620x.42b1.28.

27 Askew MJ, Steege JW, Lewis JL, Ranieri JR, Wixson RL. Effect of cement pressure and bone strength on polymethylmethacrylate fixation. *J Orthop Res* 1983. DOI:10.1002/jor.1100010410.

28 Mulroy RD, Harris WH. The effect of improved cementing techniques on component loosening in total hip replacement. An 11-year radiographic review. *J Bone Jt Surg - Ser B* 1990. DOI:10.1302/0301-620x.72b5.2211749.

29 Williams HDW, Browne G, Gie GA, Ling RSM, Timperley AJ, Wendover NA. The Exeter universal cemented femoral component at 8 to 12 years. A study of the first 325 hips. *J Bone Jt Surg - Ser B* 2002. DOI:10.1302/0301-620X.84B3.12261.

30 Hook S, Moulder E, Yates PJ, Burston BJ, Whitley E, Bannister GC. The Exeter Universal stem: a minimum ten-year review from an independent centre. *J Bone Joint Surg Br* 2006. DOI:10.1302/0301-620X.88B12.18345.

31 Garellick G, Malchau H, Regnér H, Herberts P. The Charnley versus the

Spectron hip prosthesis: Radiographic evaluation of a randomized, prospective study of 2 different hip implants. *J Arthroplasty* 1999. DOI:10.1016/S0883-5403(99)90096-7.

32 Petheram TG, Bone M, Joyce TJ, Serrano-Pedraza I, Reed MR, Partington PF. Surface finish of the Exeter Trauma Stem: A cause for concern? *J Bone Jt Surg - Ser B* 2013. DOI:10.1302/0301-620X.95B2.31001.

33 Muirhead-Allwood SK. Lessons of a hip failure. *Br Med J* 1998.

34 Harris WH, Schiller AL, Scholler JM, Freiberg RA, Scott R. Extensive localized bone resorption in the femur following total hip replacement. *J Bone Jt Surg - Ser A* 1976. DOI:10.2106/00004623-197658050-00005.

35 McAuley JP, Culpepper WJ, Engh CA, Davis L. Total hip arthroplasty: Concerns with extensively porous coated femoral components. In: Clinical Orthopaedics and Related Research. 1998. DOI:10.1097/00003086-199810000-00019.

36 Campbell ACL, Rorabeck CH, Bourne RB, Chess D, Nott L. Thigh pain after cementless hip arthroplasty: Annoyance or ill omen. *J Bone Jt Surg - Ser B* 1992. DOI:10.1302/0301-620x.74b1.1732268.

37 Lavernia C, D'Apuzzo M, Hernandez V, Lee D. Thigh pain in primary total hip arthroplasty: The effects of elastic moduli. In: Journal of Arthroplasty. 2004. DOI:10.1016/j.arth.2004.06.023.

38 Reitman RD, Emerson R, Higgins L, Head W. Thirteen year results of total hip arthroplasty using a tapered titanium femoral component inserted without cement in patients with Type C bone. In: Journal of Arthroplasty. 2003.

DOI:10.1016/S0883-5403(03)00344-9.

39 Charnley J. Low Friction Arthroplasty of the Hip: Theory and Practice. *New York Springer-Verlag* 1979.

40 Weiss RJ, Hailer NP, Stark A, Kärrholm J. Survival of uncemented acetabular monoblock cups. Acta Orthop. 2012. DOI:10.3109/17453674.2012.688726.

41 Van Praet F, Mulier M. To cement or not to cement acetabular cups in total hip arthroplasty: a systematic review and re-evaluation. *SICOT-J* 2019; **5**. https://doi.org/10.1051/sicotj/2019032.

42 Goodman SB, Gibon E, Yao Z. The basic science of periprosthetic osteolysis. *Instr Course Lect* 2013.

43 Jacobs JJ, Roebuck KA, Archibeck M, Hallab NJ, Glant TT. Osteolysis: Basic science. In: Clinical Orthopaedics and Related Research. 2001. DOI:10.1097/00003086-200112000-00008.

44 Saikko V, Calonius O, Keränen J. Wear of conventional and cross-linked ultra-high-molecular-weight polyethylene acetabular cups against polished and roughened CoCr femoral heads in a biaxial hip simulator. *J Biomed Mater Res* 2002. DOI:10.1002/jbm.10471.

45 Drummond J, Tran P, Fary C. Metal-on-Metal Hip Arthroplasty: A Review of Adverse Reactions and Patient Management. *J Funct Biomater* 2015. DOI:10.3390/jfb6030486.

46 Cohen D. Out of joint: The story of the ASR. *BMJ* 2011. DOI:10.1136/bmj.d2905.

47 Warth LC, Callaghan JJ, Liu SS, Klaassen AL, Goetz DD, Johnston RC. Thirty-five-year results after charnley total hip arthroplasty in patients less than fifty years old: A concise follow-up of previous reports. *J Bone Jt Surg - Am Vol* 2014. DOI:10.2106/JBJS.M.01573.

48 Ball ST, Le Duff MJ, Amstutz HC. Early results of conversion of a failed femoral component in hip resurfacing arthroplasty. *J Bone Jt Surg - Ser A* 2007. DOI:10.2106/JBJS.F.00708.

49 Zagra L, Gallazzi E. Bearing surfaces in primary total hip arthroplasty. *EFORT Open Rev* 2018. DOI:10.1302/2058-5241.3.180300.

50 Schmalzried TP, Shepherd EF, Dorey FJ, *et al.* The John Charnley Award. Wear is a function of use, not time. *Clin Orthop Relat Res* 2000.

51 Maloney WJ, Galante JO, Anderson M, *et al.* Fixation, polyethylene wear, and pelvic osteolysis in primary total hip replacement. In: Clinical Orthopaedics and Related Research. 1999. DOI:10.1097/00003086-199912000-00016.

52 Smith SL, Dowson D, Goldsmith AAJ. The lubrication of metal-on-metal total hip joints: A slidedown the Stribeck curve. *Proc Inst Mech Eng Part J J Eng Tribol* 2001. DOI:10.1243/1350650011543718.

53 Smith SL, Dowson D, Goldsmith AA. The effect of femoral head diameter upon lubrication and wear of metal-on-metal total hip replacements. *Proc Inst Mech Eng H* 2001; **215**: 161–70.

54 Smith SL, Dowson D, Goldsmith AAJ. The effect of diametral clearance, motion and loading cycles upon lubrication of metal-on-metal total hip replacements.

Proc Inst Mech Eng Part C J Mech Eng Sci 2001. DOI:10.1243/0954406011520454.

55 Harty JA, Devitt B, Harty LC, Molloy M, McGuinness A. Dual energy X-ray absorptiometry analysis of peri-prosthetic stress shielding in the Birmingham resurfacing hip replacement. *Arch Orthop Trauma Surg* 2005. DOI:10.1007/s00402-005-0059-4.

56 Kishida Y, Sugano N, Nishii T, Miki H, Yamaguchi K, Yoshikawa H. Preservation of the bone mineral density of the femur after surface replacement of the hip. *J Bone Jt Surg - Ser B* 2004. DOI:10.1302/0301-620X.86B2.14338.

57 Treacy RBC, McBryde CW, Pynsent PB. Birmingham hip resurfacing arthroplasty. *J Bone Jt Surg - Ser B* 2005. DOI:10.1302/0301-620X.87B2.15030.

58 Treacy RBC, McBryde CW, Shears E, Pynsent PB. Birmingham hip resurfacing: A minimum follow-up of ten years. *J Bone Jt Surg - Ser B* 2011. DOI:10.1302/0301-620X.93B1.24134.

59 Lombardi A V., Barrack RL, Berend KR, *et al.* The Hip Society: Algorithmic approach to diagnosis and management of metal-on-metal arthroplasty. J. Bone Jt. Surg. - Ser. B. 2012. DOI:10.1302/0301-620X.94B11.30680.

60 Committee M. Minutes of the Committee on the Safety of Devices Meeting 23 March 2006. .

61 Grammatopoulos G, Pandit H, Kwon YM, *et al.* Hip resurfacings revised for inflammatory pseudotumour have a poor outcome. *J Bone Jt Surg - Ser B* 2009. DOI:10.1302/0301-620X.91B8.22562.

62 Glyn-Jones S, Pandit H, Kwon YM, Doll H, Gill HS, Murray DW. Risk factors for

inflammatory pseudotumour formation following hip resurfacing. *J Bone Jt Surg - Ser B* 2009. DOI:10.1302/0301-620X.91B12.22287.

63 Korovessis P, Petsinis G, Repanti M, Repantis T. Metallosis after contemporary metal-on-metal total hip arthroplasty: Five to nine-year follow-up. J. Bone Jt. Surg. - Ser. A. 2006. DOI:10.2106/JBJS.D.02916.

64 Haddad FS, Thakrar RR, Hart AJ, *et al.* Metal-on-metal bearings: The evidence so far. J. Bone Jt. Surg. - Ser. B. 2011. DOI:10.1302/0301-620X.93B4.26429.

65 Langton DJ, Joyce TJ, Jameson SS, *et al.* Adverse reaction to metal debris following hip resurfacing: The influence of component type, orientation and volumetric wear. *J Bone Jt Surg - Ser B* 2011. DOI:10.1302/0301-620X.93B2.25099.

66 Pandit H, Vlychou M, Whitwell D, *et al.* Necrotic granulomatous pseudotumours in bilateral resurfacing hip arthoplasties: Evidence for a type IV immune response. *Virchows Arch* 2008. DOI:10.1007/s00428-008-0659-9.

67 Kwon YM, Glyn-Jones S, Simpson DJ, *et al.* Analysis of wear of retrieved metal-on-metal hip resurfacing implants revised due to pseudotumours. *J Bone Jt Surg - Ser B* 2010. DOI:10.1302/0301-620X.92B3.23281.

68 Kwon YM, Ostlere SJ, McLardy-Smith P, Athanasou NA, Gill HS, Murray DW. 'Asymptomatic' Pseudotumors After Metal-on-Metal Hip Resurfacing Arthroplasty. Prevalence and Metal Ion Study. *J Arthroplasty* 2011. DOI:10.1016/j.arth.2010.05.030.

69 Langton DJ, Jameson SS, Joyce TJ, Webb J, Nargol AVF. The effect of component size and orientation on the concentrations of metal ions after resurfacing

arthroplasty of the hip. *J Bone Jt Surg - Ser B* 2008. DOI:10.1302/0301-620X.90B9.20785.

70 De Steiger RN, Hang JR, Miller LN, Graves SE, Davidson DC. Five-year results of the ASR XL acetabular system and the ASR hip resurfacing system: An analysis from the Australian Orthopaedic Association National Joint Replacement Registry. *J Bone Jt Surg - Ser A* 2011. DOI:10.2106/JBJS.J.01727.

71 National Joint Registry. National Joint Registry for England and Wales: 10th Annual Report. 2013.

72 Langton DJ, Jameson SS, Joyce TJ, Hallab NJ, Natu S, Nargol AVF. Early failure of metal-on-metal bearings in hip resurfacing and large-diameter total hip replacement: A consequence of excess wear. *J Bone Jt Surg - Ser B* 2010. DOI:10.1302/0301-620X.92B1.22770.

73 Bolland BJRF, Culliford DJ, Langton DJ, Millington JPS, Arden NK, Latham JM. High failure rates with a large-diameter hybrid metal-on-metal total hip replacement: Clinical, radiological and retrieval analysis. *J Bone Jt Surg - Ser B* 2011. DOI:10.1302/0301-620X.93B5.26309.

74 Australian Joint Replacement Registry. Hip and Knee Arthroplasty Report. 2012.

75 Matharu GS, Matharu VK. Metal-on-metal hip replacements: Implications for general practice. *Br J Gen Pract* 2017. DOI:10.3399/bjgp17X693557.

76 Smith AJ, Dieppe P, Porter M, Blom AW. 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. DOI:10.1136/bmj.e2383.

77 Hunt LP, Blom AW, Matharu GS, Porter ML, Whitehouse MR. The risk of developing cancer following metal-on-metal hip replacement compared with non metal-on-metal hip bearings: Findings from a prospective national registry "The National Joint Registry of England, Wales, Northern Ireland and the Isle of Man". *PLoS One* 2018. DOI:10.1371/journal.pone.0204356.

78 Munro JT, Masri BA, Duncan CP, Garbuz DS. High complication rate after revision of large-head metal-on-metal total hip arthroplasty. In: Clinical Orthopaedics and Related Research. 2014. DOI:10.1007/s11999-013-2979-6.

79 Medicines and Healthcare products Regulatory Agency. Medical Device Alert, All metal-on-metal (MoM) hip replacements. *Distribution* 2010.

80 Wynn-Jones H, MacNair R, Wimhurst J, *et al.* Silent soft tissue pathology is common with a modern metal-on-metal hip arthroplasty. *Acta Orthop* 2011. DOI:10.3109/17453674.2011.579518.

81 Hart AJ, Satchithananda K, Liddle AD, *et al.* Pseudotumors in association with well-functioning metal-on-metal hip prostheses: A case-control study using three-dimensional computed tomography and magnetic resonance imaging. *J Bone Jt Surg - Ser A* 2012. DOI:10.2106/JBJS.J.01508.

82 Fehring TK, Odum S, Sproul R, Weathersbee J. High frequency of adverse local tissue reactions in asymptomatic patients with metal-on-metal THA. In: Clinical Orthopaedics and Related Research. 2014. DOI:10.1007/s11999-013-3222-1.

83 British Orthopaedic Association. BOA/BHS Statement on Metal on Metal Hip

Implants - Update from the MHRA. 2017. https://www.boa.ac.uk/resources/boa-bhs-statement-on-metal-on-metal-hip-implants-update-from-the-mhra.html (accessed Oct 8, 2019).

84 Matharu GS, Mellon SJ, Murray DW, Pandit HG. Follow-Up of Metal-on-Metal Hip Arthroplasty Patients Is Currently Not Evidence Based or Cost Effective. *J Arthroplasty* 2015. DOI:10.1016/j.arth.2015.03.009.

85 Cohen D, Billingsley M. Medical devices: Europeans are left to their own devices. BMJ. 2011. DOI:10.1136/bmj.d2748.

86 Wilmshurst P. The regulation of medical devices. *BMJ* 2011; **342**: d2822.

87 McCulloch P. The EU's system for regulating medical devices. *BMJ* 2012. DOI:10.1136/bmj.e7126.

88 Dias J, Kay P, Porter M BT. Restoring Your Mobility. British Orthopaedic Association. 2012.

89 European Commission. Commission Notice – The 'Blue Guide' on the implementation of EU products rules. 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016XC0726(02)&from=BG (accessed Oct 9, 2019).

90 Melvin T, Torre M. New medical device regulations: The regulator's view. *EFORT Open Rev* 2019. DOI:10.1302/2058-5241.4.180061.

91 European Commission. List of Notified Bodies. https://ec.europa.eu/growth/tools-databases/nando/index.cfm?fuseaction=notifiedbody.notifiedbodies&num=TA

B&text=Technical Assessment Body (accessed Oct 9, 2019).

92 Van Norman GA. Drugs and Devices: Comparison of European and U.S. Approval Processes. *JACC Basic to Transl Sci* 2016. DOI:10.1016/j.jacbts.2016.06.003.

93 Zuckerman DM, Brown P, Nissen SE. Medical device recalls and the FDA approval process. *Arch Intern Med* 2011. DOI:10.1001/archinternmed.2011.30.

94 Hauptfleisch J, Glyn-Jones S, Beard DJ, *et al.* The premature failure of the Charnley Elite-Plus stem: a confirmation of RSA predictions. *J Bone Jt Surg Br* 2006; **88**: 179–83.

95 Roy N, Hossain S, Ayeko C, McGee HM, Elsworth CF, Jacobs LGH. 3M capital hip arthroplasty. 3-8-Year follow-up of 208 primary hip replacements. *Acta Orthop Scand* 2002. DOI:10.1080/00016470216328.

96 Di Mario C, James S, Dudek D, Sabate M, Degertekin M. Commentary: The risk of over-regulation. BMJ. 2011. DOI:10.1136/bmj.d3021.

97 Johnsson R, Thorngren KG, Persson BM. Revision of total hip replacement for primary osteoarthritis. *J Bone Jt Surg - Ser B* 1988.

98 House of Commons Science and Technology Committee. Regulation of medical implants in the EU and UK: Fifth report of session 2012-13. 2012.

99 Secretary of State for Health. Government Response to the House of Commons Science & Technology Committee Report of Session 2012-13. 2012.

100 British Orthopaedic Association, Arthritis Research UK. Joint submission to the

MHRA consultation on the 'Revision of European Legislation on Medical Devices'''. 2013.

101 British Orthopaedic Association, Dias J. Project "Beyond Compliance": for the safer introduction of orthopaedic implants - statement by the BOA president, Professor Joe Dias. 2012.

102 Porter M, Armstrong R, Howard P, Porteous M, Wilkinson JM. Orthopaedic registries - the UK view (National Joint Registry): Impact on practice. *EFORT Open Rev* 2019. DOI:10.1302/2058-5241.4.180084.

103 The National Joint Registry for England and Wales. The National joint Registry 9th Annual Report. 2012.

104 Healthcare Quality Improvement Partnerships. Joint replacement surgery: the National Joint Registry. 2019. https://www.hqip.org.uk/national-programmes/joint-replacement-surgery-the-national-joint-registry/#.XaH4Ji-ZOu4 (accessed Oct 12, 2019).

105 Nelissen RGHH, Pijls BG, Kärrholm J, Malchau H, Nieuwenhuijse MJ, Valstar ER. RSA and registries: the quest for phased introduction of new implants. *J Bone Joint Surg Am* 2011. DOI:10.2106/JBJS.K.00907.

106 Versus Arthritis. Metal-on-Metal Hip Replacement Q and A. https://www.versusarthritis.org/about-arthritis/treatments/surgery/metal-on-metal-hip-replacement-q-and-a/ (accessed Oct 12, 2019).

107 NHS Supply Chain. Orthopaedic Data Evaluation Panel. http://www.odep.org.uk (accessed Oct 14, 2019).

108 Kandala NB, Connock M, Pulikottil-Jacob R, *et al.* Setting benchmark revision rates for total hip replacement: analysis of registry evidence. *BMJ* 2015; **350**: h756.

109 National Institute for Health and Clinical Excellence. Total hip replacement and resurfacing arthroplasty for end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44). NICE, 2014 https://www.nice.org.uk/guidance/ta304.

110 Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM. Evaluation and stages of surgical innovations. Lancet. 2009. DOI:10.1016/S0140-6736(09)61083-7.

111 Ergina PL, Cook JA, Blazeby JM, *et al.* Challenges in evaluating surgical innovation. Lancet. 2009. DOI:10.1016/S0140-6736(09)61086-2.

112 McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK, Group I. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* 2013; **346**: f3012.

113 McCulloch P, Altman DG, Campbell WB, *et al.* No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009. DOI:10.1016/S0140-6736(09)61116-8.

114 Graves N, Wloch C, Wilson J, *et al.* A cost-effectiveness modelling study of strategies to reduce risk of infection following primary hip replacement based on a systematic review. *Health Technol Assess (Rockv)* 2016. DOI:10.3310/hta20540.

115 Murray DW, Carr AJ, Bulstrode CJ. Which primary total hip replacement? *J Bone*

Joint Surg Br 1995; **77**: 520–7.

116 Costa M. Reforming the regulation of orthopaedic devices in the UK. *BMJ* 2012. DOI:10.1136/bmj.e7374.

117 Kynaston-Pearson F, Ashmore AM, Malak TT, *et al.* Primary hip replacement prostheses and their evidence base: systematic review of literature. *BMJ* 2013; **347**: f6956.

118 Chaverri-Fierro D, Lobo-Escolar L, Espallargues M, Martínez-Cruz O, Domingo L, Pons-Cabrafiga M. Primary total hip arthroplasty in Catalonia: What is the clinical evidence that supports our prosthesis? *Rev Española Cirugía Ortopédica y Traumatol (English Ed* 2017. DOI:10.1016/j.recote.2017.03.002.

119 National Institute for Health and Clinical Excellence. Guidance on the Selection of Prostheses for Primary Total Hip Replacement. 2000.

120 Malak TT, Broomfield JA, Palmer AJ, *et al.* Surrogate markers of long-term outcome in primary total hip arthroplasty: A systematic review. *Bone Jt Res* 2016; **5**: 206–14.

121 Cohen D. How a fake hip showed up failings in European device regulation. *BMJ* 2012. DOI:10.1136/bmj.e7090.

122 Lübbeke A, Silman AJ, Prieto-Alhambra D, Adler AI, Barea C, Carr AJ. The role of national registries in improving patient safety for hip and knee replacements. BMC Musculoskelet. Disord. 2017. DOI:10.1186/s12891-017-1773-0.

123 Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG, Group I. IDEAL framework for surgical innovation 2: observational studies in the exploration

and assessment stages. *BMJ* 2013; **346**: f3011.

124 Marinac-Dabic D, Normand SL, Sedrakyan A, Gross T. Epidemiologic Studies of Medical Devices: Methodologic Considerations for Implantable Devices. In: Pharmacoepidemiology, Fifth Edition. 2012. DOI:10.1002/9781119959946.ch27.

125 Sayers A, Crowther MJ, Judge A, Whitehouse MR, Blom AW. Determining the sample size required to establish whether a medical device is non-inferior to an external benchmark. *BMJ Open* 2017. DOI:10.1136/bmjopen-2016-015397.

126 Orthopaedic Data Evaluation Panel. ODEP Benchmark Developments 2017. 2017. http://www.odep.org.uk/Portals/0/ODEPStatements/ODEP Benchmarks 2017_rev.6.pdf (accessed May 13, 2020).

127 Pane J, Francisca RDC, Verhamme KMC, *et al.* EU postmarket surveillance plans for medical devices. *Pharmacoepidemiol Drug Saf* 2019. DOI:10.1002/pds.4859.

128 Feigal DW. Postmarket Considerations in Biomaterials and Medical Devices. In: Biomaterials Science: An Introduction to Materials: Third Edition. 2013. DOI:10.1016/B978-0-08-087780-8.00139-X.

129 Ghogomu EAT, Maxwell LJ, Buchbinder R, *et al.* Updated method guidelines for cochrane musculoskeletal group systematic reviews and metaanalyses. J. Rheumatol. 2014. DOI:10.3899/jrheum.121306.

130 National Institute for Health and Clinical Excellence. Total hip replacement and resurfacing arthroplasty for end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44). 2014 http://www.nice.org.uk/guidance/ta304.

131 Lenguerrand E, Whitehouse MR, Beswick AD, *et al.* Risk factors associated with

revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. *Lancet Infect Dis* 2018. DOI:10.1016/S1473-3099(18)30345-1.

132 Prokopetz JJ, Losina E, Bliss RL, Wright J, Baron JA, Katz JN. Risk factors for revision of primary total hip arthroplasty: a systematic review. *BMC Musculoskelet Disord* 2012; **13**: 251.

133 McKee AE, Farrell AT, Pazdur R, Woodcock J. The Role of the U.S. Food and Drug Administration Review Process: Clinical Trial Endpoints in Oncology. *Oncologist* 2010. DOI:10.1634/theoncologist.2010-s1-13.

134 Sesselmann S, Forst R, Tschunko F. Radiostereometric analysis of hip implants: A critical review of methodology and future directions. *OA Musculoskelet Med* 2013. DOI:10.13172/2052-9287-1-4-941.

135 Pijls BG, Nieuwenhuijse MJ, Fiocco M, *et al.* Early proximal migration of cups is associated with late revision in THA: A systematic review and meta-analysis of 26 RSA studies and 49 survival studies. *Acta Orthop* 2012. DOI:10.3109/17453674.2012.745353.

136 Karrholm J, Borssen B, Lowenhielm G, Snorrason F. Does early micromotion of femoral stem prostheses matter? 4-7-year stereoradiographic follow-up of 84 cemented prostheses. *J Bone Jt Surg - Ser B* 1994. DOI:10.1302/0301-620x.76b6.7983118.

137 Freeman MAR, Plante-Bordeneuve P. Early migration and late aseptic failure of proximal femoral prostheses. *J Bone Jt Surg - Ser B* 1994. DOI:10.1302/0301-620x.76b3.8175848.

138 Nieuwenhuijse MJ, Valstar ER, Kaptein BL, Nelissen RGHH. Good diagnostic performance of early migration as a predictor of late aseptic loosening of acetabular cups: Results from ten years of follow-up with Roentgen Stereophotogrammetric Analysis (RSA). *J Bone Jt Surg - Ser A* 2012. DOI:10.2106/JBJS.K.00305.

139 Krismer M, Stöckl B, Fischer M, Bauer R, Mayrhofer P, Ogon M. Early migration predicts late aseptic failure of hip sockets. *J Bone Jt Surg - Ser B* 1996. DOI:10.1302/0301-620x.78b3.0780422.

140 Hendrich C, Ak D, Sauer U, Kirschner S, Schmitz H, Martell JM. High long-term loosening rate of conical screw cups. *Acta Orthop* 2006. DOI:10.1080/17453670610013178.

141 Krismer M, Biedermann R, Stöckl B, Fischer M, Bauer R, Haid C. The prediction of failure of the stem in THR by measurement of early migration using EBRA-FCA. Einzel-Bild-Roentgen-Analyse-femoral component analysis. *J Bone Joint Surg Br* 1999.

142 Mazoochian F, Schrimpf FM, Kircher J, *et al.* Proximal loading of the femur leads to low subsidence rates: First clinical results of the CR-stem. *Arch Orthop Trauma Surg* 2007. DOI:10.1007/s00402-007-0384-x.

143 Takenaga RK, Callaghan JJ, Bedard NA, Liu SS, Gao Y. Which functional assessments predict long-term wear after total hip arthroplasty? *Clin Orthop Relat Res* 2013. DOI:10.1007/s11999-013-2968-9.

144 Kobayashi A, Donnelly WJ, Scott G, Freeman MAR. Early radiological observations may predict the long-term survival of femoral hip prostheses. In:

Journal of Bone and Joint Surgery - Series B. 1997. DOI:10.1302/0301-620X.79B4.7210.

145 Ranawat CS, Deshmukh RG, Peters LE, Umlas ME. Prediction of the long-term durability of all-polyethylene cemented sockets. In: Clinical Orthopaedics and Related Research. 1995.

146 Dowd JE, Sychterz CJ, Young AM, Engh CA. Characterization of long-term femoral-head-penetration rates: Association with and prediction of osteolysis. *J Bone Jt Surg - Ser A* 2000. DOI:10.2106/00004623-200008000-00006.

147 Malik MHA, Fisher N, Gray J, Wroblewski BM, Kay PR. Prediction of Charnley femoral stem aseptic loosening by early post-operative radiological features. *Int Orthop* 2005. DOI:10.1007/s00264-005-0667-1.

148 Scott G, Nakagawa S, Orhan Z, Freeman M. The shortcomings of computer-aided measurement of migration for the prediction of failure of three forms of acetabular fixation by survival analysis and migration study to ten years. *HIP Int* 2006. DOI:10.5301/HIP.2008.1277.

149 Khalily C, Whiteside LA. Predictive value of early radiographic findings in cementless total hip arthroplasty femoral components: An 8- to 12-year follow-up. *J Arthroplasty* 1998. DOI:10.1016/S0883-5403(98)90028-6.

150 Nieuwenhuijse MJ, Valstar ER, Kaptein BL, Nelissen RG. The Exeter femoral stem continues to migrate during its first decade after implantation: 10-12 years of follow-up with radiostereometric analysis (RSA). *Acta Orthop* 2012; **83**: 129–34.

151 Harris WH. Traumatic arthritis of the hip after dislocation and acetabular

fractures: treatment by mold arthroplasty. *J Bone Jt Surg Am* 1969.

152 Pilliar RM, Lee JM, Maniatopoulos C. Observations on the effect of movement on bone ingrowth into porous-surfaced implants. *Clin Orthop Relat Res* 1986. DOI:10.1097/00003086-198607000-00023.

153 Soballe K, Hansen ES, Brockstedt-Rasmussen H, Bunger C. Hydroxyapatite coating converts fibrous tissue to bone around loaded implants. *J Bone Jt Surg - Ser B* 1993. DOI:10.1302/0301-620x.75b2.8444949.

154 Krismer M, Bauer R, Tschupik J, Mayrhofer P. EBRA: A method to measure migration of acetabular components. *J Biomech* 1995. DOI:10.1016/0021-9290(94)00177-6.

155 Dumbleton JH, Manley MT, Edidin AA. A literature review of the association between wear rate and osteolysis in total hip arthroplasty. *J Arthroplasty* 2002. DOI:10.1054/arth.2002.33664.

156 Davidson JM. ROENTGEN RAYS and LOCALISATION: An Apparatus for Exact Measurement and Localisation by Means of Roentgen Rays. *BMJ* 1898. DOI:10.1136/bmj.1.1931.10.

157 Valstar ER, Gill R, Ryd L, Flivik G, Börlin N, Kärrholm J. Guidelines for standardization of radiostereometry (RSA) of implants. *Acta Orthop* 2005. DOI:10.1080/17453670510041574.

158 Thomas GER, Simpson DJ, Mehmood S, *et al.* The seven-year wear of highly cross-linked polyethylene in total hip arthroplasty: A double-blind, randomized controlled trial using radiostereometric analysis. *J Bone Jt Surg - Ser A* 2011.

DOI:10.2106/JBJS.J.00287.

159 Mäkinen TJ, Koort JK, Mattila KT, Aro HT. Precision measurements of the RSA method using a phantom model of hip prosthesis. *J Biomech* 2004. DOI:10.1016/j.jbiomech.2003.09.004.

160 Bottner F, Su E, Nestor B, Azzis B, Sculco TP, Bostrom M. Radiostereometric Analysis: The Hip. *HSS J* 2005. DOI:10.1007/s11420-005-0114-2.

161 Malchau H. Introducing new technology: A stepwise algorithm. Spine (Phila. Pa. 1976). 2000. DOI:10.1097/00007632-200002010-00004.

162 Valstar ER, Gill R. Radiostereometric analysis in orthopaedic surgery: Editorial comment. In: Clinical Orthopaedics and Related Research. 2006. DOI:10.1097/01.blo.0000224017.80787.7f.

163 Hurschler C, Seehaus F, Emmerich J, Kaptein BL, Windhagen H. Comparison of the Model-Based and Marker-Based Roentgen Stereophotogrammetry Methods in a Typical Clinical Setting. *J Arthroplasty* 2009. DOI:10.1016/j.arth.2008.02.004.

164 Frazer AR, Tanzer M. New Implant Introduction in Total Hip Arthroplasty Using Radiostereometric Analysis: A Cautionary Note. *J Arthroplasty* 2019. DOI:10.1016/j.arth.2019.10.004.

165 de Vries LMA, van der Weegen W, Pilot P, Stolarczyk PA, Sijbesma T, Hoffman EL. The predictive value of radiostereometric analysis for stem survival in total hip arthroplasty. A systematic review. HIP Int. 2014. DOI:10.5301/hipint.5000102.

166 Kärrholm J, Gill RHS, Valstar ER. The history and future of radiostereometric

analysis. In: Clinical Orthopaedics and Related Research. 2006. DOI:10.1097/01.blo.0000224001.95141.fe.

167 Kärrholm J, Herberts P, Hultmark P, Malchau H, Nivbrant B, Thanner J. Radiostereometry of hip prostheses: Review of methodology and clinical results. In: Clinical Orthopaedics and Related Research. 1997. DOI:10.1097/00003086-199711000-00011.

168 Derbyshire B, Prescott RJ, Porter ML. Notes on the use and interpretation of radiostereometric analysis. *Acta Orthop* 2009. DOI:10.1080/17453670902807474.

169 Salemyr M, Muren O, Ahl T, *et al.* Lower periprosthetic bone loss and good fixation of an ultra-short stem compared to a conventional stem in uncemented total hip arthroplasty. *Acta Orthop* 2015. DOI:10.3109/17453674.2015.1067087.

170 Australian Orthopaedic Association National Joint Replacement Registry. Annual Report. 2016.

171 Weber E, Sundberg M, Flivik G. Design modifications of the uncemented Furlong hip stem result in minor early subsidence but do not affect further stability: A randomized controlled RSA study with 5-year follow-up. *Acta Orthop* 2014. DOI:10.3109/17453674.2014.958810.

172 Baad-Hansen T, Kold S, Olsen N, Christensen F, Søballe K. Excessive distal migration of fiber-mesh coated femoral stems. *Acta Orthop* 2011. DOI:10.3109/17453674.2011.574562.

173 Penny JO, Ding M, Varmarken JE, Ovesen O, Overgaard S. Early micromovement of the Articular Surface Replacement (ASR) femoral

component: Two-year radiostereometry results. *J Bone Jt Surg - Ser B* 2012. DOI:10.1302/0301-620X.94B10.29030.

174 Rothwell AG, Hooper GJ, Hobbs A, Frampton CM. An analysis of the Oxford hip and knee scores and their relationship to early joint revision in the New Zealand Joint Registry. *J Bone Jt Surg Br* 2010; **92**: 413–8.

175 Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. *BMJ* 2015. DOI:10.1136/bmj.g7818.

176 Hailer NP, Garellick G, Kärrholm J. Uncemented and cemented primary total hip arthroplasty in the Swedish Hip Arthroplasty Register: Evaluation of 170,413 operations. *Acta Orthop* 2010. DOI:10.3109/17453671003685400.

177 Cohen D. Out of joint: the story of the ASR. *BMJ* 2011; **342**: d2905.

178 Havelin LI, Espehaug B, Vollset SE, Engesaeter LB. The effect of the type of cement on early revision of Charnley total hip prostheses. A review of eight thousand five hundred and seventy-nine primary arthroplasties from the Norwegian Arthroplasty Register. *J Bone Jt Surg Am* 1995; **77**: 1543–50.

179 Muirhead-Allwood SK. Lessons of a hip failure. *BMJ* 1998; **316**: 644.

180 McCulloch P, Altman DG, Campbell WB, *et al.* No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009; **374**: 1105–12.

181 Nelissen RG, Pijls BG, Karrholm J, Malchau H, Nieuwenhuijse MJ, Valstar ER. RSA and registries: the quest for phased introduction of new implants. *J Bone Jt Surg Am* 2011; **93 Suppl 3**: 62–5.

182 Sesselmann S, Forst R, Tschunko F. Radiostereometric analysis of hip implants: a critical review of methodology and future directions. *OA Musculoskelet Med* 2013; : 4.

183 Penny JO, Ding M, Varmarken JE, Ovesen O, Overgaard S. Early micromovement of the Articular Surface Replacement (ASR) femoral component: two-year radiostereometry results. *J Bone Jt Surg Br* 2012; **94**: 1344–50.

184 McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK, IDEAL Group. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* 2013. DOI:10.1136/bmj.f3012.

185 European Commission. Press Release - Medical devices: European Commission calls for immediate actions. 2012. https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_12_77 (accessed March 18, 2020).

186 REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EE. 2017. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745 (accessed March 20, 2020).

187 REGULATION (EU) 2017/746 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. 2017.

http://www.ce-mark.com/IVD Regulation.pdf (accessed March 20, 2020).

188 UK Government. Regulating medical devices in the UK. 2020. https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk (accessed Jan 12, 2021).

189 Labek G, Neumann D, Agreiter M, Schuh R, Böhler N. Impact of implant developers on published outcome and reproducibility of cohort-based clinical studies in arthroplasty. *J Bone Joint Surg Am* 2011. DOI:10.2106/JBJS.K.01108.

190 Labek G, Frischhut S, Schlichtherle R, Williams A, Thaler M. Outcome of the cementless Taperloc stem: A comprehensive literature review including arthroplasty register data. *Acta Orthop* 2011. DOI:10.3109/17453674.2011.570668.

191 Sedrakyan A, Paxton EW, Phillips C, *et al.* The International Consortium of Orthopaedic Registries: overview and summary. J. Bone Joint Surg. Am. 2011. DOI:10.2106/jbjs.k.01125.

192 Horan FT. Joint registries. J. Bone Jt. Surg. - Ser. B. 2010. DOI:10.1302/0301-620X.92B6.24549.

193 Knutson K, Robertsson O. The Swedish Knee Arthroplasty Register (www.knee.se): The inside story. Acta Orthop. 2010. DOI:10.3109/17453671003667267.

194 Sedrakyan A, Normand SLT, 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. DOI:10.1136/bmj.d7434.

195 Goodfellow JW, O'Connor JJ, Murray DW. A critique of revision rate as an

outcome measure: Re-interpretation of knee joint registry data. J. Bone Jt. Surg. - Ser. B. 2010. DOI:10.1302/0301-620X.92B12.25193.

196 Lübbeke A, Katz JN, Perneger T V., Hoffmeyer P. Primary and revision hip arthroplasty: 5-Year outcomes and influence of age and comorbidity. *J Rheumatol* 2007.

197 Espehaug B, Havelin LI, Engesæter LB, Langeland N, Vollset SE. Patient satisfaction and function after primary and revision total hip replacement. In: Clinical Orthopaedics and Related Research. 1998. DOI:10.1097/00003086-199806000-00017.

198 Adelani MA, Crook K, Barrack RL, Maloney WJ, Clohisy JC. What is the prognosis of revision total hip arthroplasty in patients 55 years and younger? *Clin Orthop Relat Res* 2014. DOI:10.1007/s11999-013-3377-9.

199 Lübbeke A, Gonzalez A, Garavaglia G, *et al.* A comparative assessment of small-head metal-on-metal and ceramic-on-polyethylene total hip replacement. *Bone Jt J* 2014. DOI:10.1302/0301-620X.96B7.32812.

200 Norton MR, Yarlagadda R, Anderson GH. Catastrophic failure of the Elite Plus total hip replacement, with a Hylamer acetabulum and Zirconia ceramic femoral head. *J Bone Jt Surg - Ser B* 2002. DOI:10.1302/0301-620X.84B5.12679.

201 England. TRC of S of. An Investigation of the Performance of the 3M™ Capital™ Hip System. 2001. https://www.rcseng.ac.uk/surgeons/research/surgical-research/docs/An Investigation of the Performance of the 3M Capital Hip System Report 2001.pdf/view.

202 Wylde V, Blom AW. The failure of survivorship. J. Bone Jt. Surg. - Ser. B. 2011. DOI:10.1302/0301-620X.93B5.26687.

203 Scott CEH, Howie CR, MacDonald D, Biant LC. Predicting dissatisfaction following total knee replacement: A prospective study of 1217 patients. *J Bone Jt Surg - Ser B* 2010. DOI:10.1302/0301-620X.92B9.24394.

204 Murray DW, Frost SJD. Pain in the assessment of total knee replacement. *J Bone Jt Surg - Ser B* 1998. DOI:10.1302/0301-620X.80B3.7820.

205 Bozic KJ, Kurtz S, Lau E, *et al.* The epidemiology of bearing surface usage in total hip arthroplasty in the United States. *J Bone Jt Surg - Ser A* 2009. DOI:10.2106/JBJS.H.01220.

206 McCulloch P, Barkun J, Sedrakyan A. Implantable device regulation in Europe. Lancet. 2012. DOI:10.1016/S0140-6736(12)61405-6.

207 Malchau H, Graves SE, Porter M, Harris WH, Troelsen A. The next critical role of orthopedic registries. Acta Orthop. 2015. DOI:10.3109/17453674.2014.1002184.

208 Avery AJ, Anderson C, Bond CM, *et al.* Evaluation of patient reporting of adverse drug reactions to the UK 'Yellow card scheme': Literature review, descriptive and qualitative analyses, and questionnaire surveys. Health Technol. Assess. (Rockv). 2011. DOI:10.3310/hta15200.

209 Faraoni D, Schaefer ST. Randomized controlled trials vs. observational studies: Why not just live together? BMC Anesthesiol. 2016. DOI:10.1186/s12871-016-0265-3.

210 Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: A

critical comparison of trials with nonrandomized studies. Hepatology. 2006. DOI:10.1002/hep.21404.

211 Gordon M, Rysinska A, Garland A, *et al.* Increased Long-Term Cardiovascular Risk after Total Hip Arthroplasty. *Med (United States)* 2016. DOI:10.1097/MD.0000000000002662.

212 Robertson O, Stefánsdóttir A, Lidgren L, Ranstam J. Increased long-term mortality in patient less than 55 years old who have undergone knee replacement for osteoarthritis. Results from the Swedish knee arthroplasty register. *J Bone Jt Surg - Ser B* 2007. DOI:10.1302/0301-620X.89B5.18355.

213 Pane J, Verhamme KMC, Rebollo I, Sturkenboom MCJM. Descriptive analysis of postmarket surveillance data for hip implants. *Pharmacoepidemiol Drug Saf* 2020. DOI:10.1002/pds.4971.

214 Marcus HJ, Payne CJ, Hughes-Hallett A, *et al.* Regulatory approval of new medical devices: Cross sectional study. *BMJ* 2016. DOI:10.1136/bmj.i2587.

215 Field RE, Cronin MD, Singh PJ. The Oxford hip scores for primary and revision hip replacement. *J Bone Jt Surg - Ser B* 2005. DOI:10.1302/0301-620X.87B5.15390.

216 Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Jt Surg - Ser B* 1996. DOI:10.1302/0301-620x.78b2.0780185.

217 Harris K, Dawson J, Gibbons E, *et al.* Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas* 2016.

DOI:10.2147/prom.s97774.

218 Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Jt Surg - Ser B* 2001. DOI:10.1302/0301-620X.83B8.11643.

219 Wylde V V., Learmonth ID, Cavendish VJ. The Oxford hip score: The patient's perspective. *Health Qual Life Outcomes* 2005. DOI:10.1186/1477-7525-3-66.

220 NHS. Finalised Patient Reported Outcome Measures (PROMs) in England for Hip & Knee Replacements, April 2018 – March 2019. 2019 https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/finalised-hip--knee-replacements-april-2018---march-2019.

221 Eneqvist T, Nemes S, Bülow E, Mohaddes M, Rolfson O. Can patient-reported outcomes predict re-operations after total hip replacement? *Int Orthop* 2018. DOI:10.1007/s00264-017-3711-z.

222 Robertsson O, Dunbar MJ. Patient satisfaction compared with general health and disease-specific questionnaires in knee arthroplasty patients. *J Arthroplasty* 2001. DOI:10.1054/arth.2001.22395a.

223 Malchau H, Garellick G, Eisler T, Kärrholm J, Herberts P. Presidential guest address: The Swedish Hip Registry: Increasing the sensitivity by patient outcome data. In: Clinical Orthopaedics and Related Research. 2005. DOI:10.1097/01.blo.0000193517.19556.e4.

224 Dy CJ, Bozic KJ, Pan TJ, Wright TM, Padgett DE, Lyman S. Risk factors for early revision after total hip arthroplasty. *Arthritis Care Res* 2014; **66**: 907–15.

225 Fitzpatrick R, Morris R, Hajat S, *et al.* The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. *Qual Saf Heal Care* 2000. DOI:10.1136/qhc.9.3.146.

226 Streiner DL, Norman GR. Health Measurement Scales: A practical guide to their development and use. 2008 DOI:10.1093/acprof:oso/9780199231881.001.0001.

227 McMurray R, Heaton J, Sloper P, Nettleton S. Measurement of patient perceptions of pain and disability in relation to total hip replacement: The place of the Oxford hip score in mixed methods. *Qual Heal Care* 1999. DOI:10.1136/qshc.8.4.228.

228 Ostendorf M, van Stel HF, Buskens E, *et al.* Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Jt Surg - Ser B* 2004. DOI:10.1302/0301-620X.86B6.14950.

229 Fitzpatrick R, Dawson J. Health-related quality of life and the assessment of outcomes of total hip replacement surgery. *Psychol Heal* 1997. DOI:10.1080/08870449708406740.

230 Wylde V, Blom AW, Whitehouse SL, Taylor AH, Pattison GT, Bannister GC. Patient-Reported Outcomes After Total Hip and Knee Arthroplasty. Comparison of Midterm Results. *J Arthroplasty* 2009. DOI:10.1016/j.arth.2007.12.001.

231 Lübbeke A, Silman AJ, Barea C, Prieto-Alhambra D, Carr AJ. Mapping existing hip and knee replacement registries in Europe. Health Policy (New. York). 2018.

DOI:10.1016/j.healthpol.2018.03.010.

232 Dimitroulas T, Duarte R V., Behura A, Kitas GD, Raphael JH. Neuropathic pain in osteoarthritis: A review of pathophysiological mechanisms and implications for treatment. Semin. Arthritis Rheum. 2014. DOI:10.1016/j.semarthrit.2014.05.011.

233 Sharma L, Kapoor D, Issa S. Epidemiology of osteoarthritis: An update. Curr. Opin. Rheumatol. 2006. DOI:10.1097/01.bor.0000209426.84775.f8.

234 Creamer P. Osteoarthritis pain and its treatment. Curr. Opin. Rheumatol. 2000. DOI:10.1097/00002281-200009000-00019.

235 Glyn-Jones S, Palmer AJR, Agricola R, *et al.* Osteoarthritis. In: The Lancet. 2015. DOI:10.1016/S0140-6736(14)60802-3.

236 Namba RS, Paxton EW, Inacio MC. Opioid Prescribers to Total Joint Arthroplasty Patients Before and After Surgery: The Majority Are Not Orthopedists. *J Arthroplasty* 2018. DOI:10.1016/j.arth.2018.05.034.

237 Bijlsma JWJ, Berenbaum F, Lafeber FPJG. Osteoarthritis: An update with relevance for clinical practice. Lancet. 2011. DOI:10.1016/S0140-6736(11)60243-2.

238 Pivec R, Johnson AJ, Mears SC, Mont MA. Hip arthroplasty. In: The Lancet. 2012. DOI:10.1016/S0140-6736(12)60607-2.

239 Darzi A. High quality care for all: NHS next stage review final report. 2008 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf.

240 Gademan MGJ, Hofstede SN, Vliet Vlieland TPM, Nelissen RGHH, Marang-Van de Mheen PJ. Indication criteria for total hip or knee arthroplasty in osteoarthritis: a state-of-the-science overview. *BMC Musculoskelet Disord* 2016. DOI:10.1186/s12891-016-1325-z.

241 Dreinhöfer KE, Dieppe P, Stürmer T, *et al.* Indications for total hip replacement: Comparison of assessments of orthopaedic surgeons and referring physicians. *Ann Rheum Dis* 2006. DOI:10.1136/ard.2005.047811.

242 Crawford RW, Murray DW. Total hip replacement: Indications for surgery and risk factors for failure. Ann. Rheum. Dis. 1997. DOI:10.1136/ard.56.8.455.

243 Maria C, Inacio M, Pratt NL, Roughead EE, Paxton EW, Graves SE. Opioid use after total hip arthroplasty surgery is associated with revision surgery. *Bmc Musculoskelet Disord* 2016; **17**: no pagination.

244 Zhang W, Moskowitz RW, Nuki G, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthr Cartil* 2008; **16**: 137–62.

245 Alviar MJ, Olver J, Brand C, *et al.* Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. J. Rehabil. Med. 2011. DOI:10.2340/16501977-0828.

246 Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013. DOI:10.1136/bmj.f167.

247 Halawi MJ. Outcome measures in total joint arthroplasty: Current status, challenges, and future directions. *Orthopedics* 2015. DOI:10.3928/01477447-

20150804-55.

248 Arias-de la Torre J, Puigdomenech E, Valderas JM, *et al.* Availability of specific tools to assess patient reported outcomes in hip arthroplasty in Spain. Identifying the best candidates to incorporate in an arthroplasty register. A systematic review and standardized assessment. *PLoS One* 2019. DOI:10.1371/journal.pone.0214746.

249 Ekman EF, Koman LA. Acute pain following musculoskeletal injuries and orthopaedic surgery: mechanisms and management. *Instr Course Lect* 2005; **54**: 21–33.

250 Nikolajsen L, Brandsborg B, Lucht U, Jensen TS, Kehlet H. Chronic pain following total hip arthroplasty: A nationwide questionnaire study. *Acta Anaesthesiol Scand* 2006. DOI:10.1111/j.1399-6576.2006.00976.x.

251 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. DOI:10.1136/bmjopen-2011-000435.

252 Sadoghi P, Liebensteiner M, Agreiter M, Leithner A, Bohler N, Labek G. Revision surgery after total joint arthroplasty: a complication-based analysis using worldwide arthroplasty registers. *J Arthroplast* 2013; **28**: 1329–32.

253 Alviar MJ, Olver J, Brand C, Hale T, Khan F. Do patient-reported outcome measures used in assessing outcomes in rehabilitation after hip and knee arthroplasty capture issues relevant to patients? Results of a systematic review and ICF linking process. *J Rehabil Med* 2011; **43**: 374–81.

254 Matharu GS, Judge A, Pandit HG, Murray DW. Which factors influence the rate of failure following metal-on-metal hip arthroplasty revision surgery performed for adverse reactions to metal debris? *Bone Jt J* 2017. DOI:10.1302/0301-620X.99B8.BJJ-2016-0889.R1.

255 Britton AR, Murray DW, Bulstrode CJ, McPherson K, Denham RA. Pain levels after total hip replacement: Their use as endpoints for survival analysis. *J Bone Jt Surg - Ser B* 1997. DOI:10.1302/0301-620X.79B1.6572.

256 Aghayev E, Teuscher R, Neukamp M, *et al.* The course of radiographic loosening, pain and functional outcome around the first revision of a total hip arthroplasty. *BMC Musculoskelet Disord* 2013. DOI:10.1186/1471-2474-14-167.

257 Malfait AM, Schnitzer TJ. Towards a mechanism-based approach to pain management in osteoarthritis. Nat. Rev. Rheumatol. 2013. DOI:10.1038/nrrheum.2013.138.

258 Namba RS, Inacio MC, Pratt NL, *et al.* Postoperative opioid use as an early indication of total hip arthroplasty failure. *Acta Orthop* 2016; **87 Suppl 1**: 37–43.

259 Blågestad T, Nordhus IH, Grønli J, *et al.* Prescription trajectories and effect of total hip arthroplasty on the use of analgesics, hypnotics, antidepressants, and anxiolytics: Results from a population of total hip arthroplasty patients. *Pain* 2016. DOI:10.1097/j.pain.0000000000000414.

260 Kolasinski SL, Neogi T, Hochberg MC, *et al.* 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. *Arthritis Rheumatol* 2020. DOI:10.1002/art.41142.

261 Kingsbury SR, Gross HJ, Isherwood G, Conaghan PG. Osteoarthritis in europe: Impact on health status, work productivity and use of pharmacotherapies in five European countries. *Rheumatol (United Kingdom)* 2014. DOI:10.1093/rheumatology/ket463.

262 Wolfe F, Zhao S, Lane N. Preference for nonsteroidal antiinflammatory drugs over acetaminophen by rheumatic disease patients: A survey of 1,799 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. *Arthritis Rheum* 2000. DOI:10.1002/1529-0131(200002)43:2<378::AID-ANR18>3.0.CO;2-2.

263 Ong CKS, Lirk P, Tan CH, Seymour RA. An evidence-based update on nonsteroidal anti-inflammatory drugs. Clin. Med. Res. 2007. DOI:10.3121/cmr.2007.698.

264 Zhang W, Doherty M, Arden N, *et al.* EULAR evidence based recommendations for the management of hip osteoarthritis: Report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 2005. DOI:10.1136/ard.2004.028886.

265 Espehaug B, Havelin LI, Engesaeter LB, Langeland N, Vollset SE. Patient-related risk factors for early revision of total hip replacements. A population register-based case-control study of 674 revised hips. *Acta Orthop Scand* 1997; **68**: 207–15.

266 Rajamäki TJ, Puolakka PA, Hietaharju A, Moilanen T, Jämsen E. Use of prescription analgesic drugs before and after hip or knee replacement in patients with osteoarthritis. *BMC Musculoskelet Disord* 2019. DOI:10.1186/s12891-019-2809-4.

267 Colvin LA, Bull F, Hales TG. Perioperative opioid analgesia—when is enough

too much? A review of opioid-induced tolerance and hyperalgesia. Lancet. 2019. DOI:10.1016/S0140-6736(19)30430-1.

268 Kehlet H, Dahl JB. Anaesthesia, surgery, and challenges in postoperative recovery. Lancet. 2003. DOI:10.1016/S0140-6736(03)14966-5.

269 Bergman S. Management of musculoskeletal pain. Best Pract. Res. Clin. Rheumatol. 2007. DOI:10.1016/j.berh.2006.10.001.

270 Kim SC, Choudhry N, Franklin JM, *et al.* Patterns and predictors of persistent opioid use following hip or knee arthroplasty. *Osteoarthr Cartil* 2017. DOI:10.1016/j.joca.2017.04.002.

271 van de Laar M. Pain Treatment in Arthritis-Related Pain: Beyond NSAIDs. *Open Rheumatol J* 2012. DOI:10.2174/1874312901206010320.

272 Zarling BJ, Yokhana SS, Herzog DT, Markel DC. Preoperative and Postoperative Opiate Use by the Arthroplasty Patient. *J Arthroplasty* 2016. DOI:10.1016/j.arth.2016.03.061.

273 da Costa BR, Nüesch E, Kasteler R, *et al.* Oral or transdermal opioids for osteoarthritis of the knee or hip. Cochrane Database Syst. Rev. 2014. DOI:10.1002/14651858.CD003115.pub4.

274 Volkow ND, McLellan TA, Cotto JH, Karithanom M, Weiss SRB. Characteristics of opioid prescriptions in 2009. JAMA - J. Am. Med. Assoc. 2011. DOI:10.1001/jama.2011.401.

275 World Health Organisation (WHO). WHO's pain ladder. World Heal. Organ. 2012.

276 Raffa RB, Friderichs E. The basic science aspect of tramadol hydrochloride. Pain Rev. 1996.

277 Busse JW, Wang L, Kamaleldin M, *et al.* Opioids for Chronic Noncancer Pain: A Systematic Review and Meta-analysis. JAMA - J. Am. Med. Assoc. 2018. DOI:10.1001/jama.2018.18472.

278 April KT, Bisaillon J, Welch V, *et al.* Tramadol for osteoarthritis. Cochrane Database Syst. Rev. 2019. DOI:10.1002/14651858.CD005522.pub3.

279 Adams EH, Breiner S, Cicero TJ, *et al.* A Comparison of the Abuse Liability of Tramadol, NSAIDs, and Hydrocodone in Patients with Chronic Pain. *J Pain Symptom Manage* 2006. DOI:10.1016/j.jpainsymman.2005.10.006.

280 Britton AR, Murray DW, Bulstrode CJ, McPherson K, Denham RA. Pain levels after total hip replacement: their use as endpoints for survival analysis. *J Bone Jt Surg Br* 1997; **79**: 93–8.

281 Dray A. New Horizons in Pharmacologic Treatment for Rheumatic Disease Pain. Rheum. Dis. Clin. North Am. 2008. DOI:10.1016/j.rdc.2008.04.005.

282 Phillips K, Clauw DJ. Central pain mechanisms in the rheumatic diseases: Future directions. Arthritis Rheum. 2013. DOI:10.1002/art.37739.

283 Kidd BL. Osteoarthritis and joint pain. Pain. 2006. DOI:10.1016/j.pain.2006.04.009.

284 Schaible HG. Mechanisms of chronic pain in osteoarthritis. Curr. Rheumatol. Rep. 2012. DOI:10.1007/s11926-012-0279-x.

285 Shigemura T, Ohtori S, Kishida S, *et al.* Neuropathic pain in patients with osteoarthritis of hip joint. *Eur Orthop Traumatol* 2011. DOI:10.1007/s12570-011-0070-x.

286 Valdes AM, Doherty SA, Zhang W, Muir KR, Maciewicz RA, Doherty M. Inverse relationship between preoperative radiographic severity and postoperative pain in patients with osteoarthritis who have undergone total joint arthroplasty. *Semin Arthritis Rheum* 2012. DOI:10.1016/j.semarthrit.2011.07.002.

287 Hochman JR, French MR, Bermingham SL, Hawker GA. The nerve of osteoarthritis pain. *Arthritis Care Res* 2010. DOI:10.1002/acr.20142.

288 Aşkın A, Özkan A, Tosun A, Demirdal ÜS, İsnaç F. Quality of life and functional capacity are adversely affected in osteoarthritis patients with neuropathic pain. *Kaohsiung J Med Sci* 2017. DOI:10.1016/j.kjms.2016.12.007.

289 Blikman T, Rienstra W, Van Raay JJAM, *et al.* Neuropathic-like symptoms and the association with joint-specific function and quality of life in patients with hip and knee osteoarthritis. *PLoS One* 2018. DOI:10.1371/journal.pone.0199165.

290 Kersten P, White PJ, Tennant A, Perrot S, Tick H. Targeting Pain or Osteoarthritis? Implications. *PAIN Clin Updat* 2011. DOI:10.1093/ecam/nep141.

291 Fu K, Robbins SR, McDougall JJ. Osteoarthritis: The genesis of pain. *Rheumatol (United Kingdom)* 2018. DOI:10.1093/rheumatology/kex419.

292 Arendt-Nielsen L. Pain sensitisation in osteoarthritis. *Clin Exp Rheumatol* 2017.

293 Thakur M, Dickenson AH, Baron R. Osteoarthritis pain: Nociceptive or neuropathic? Nat. Rev. Rheumatol. 2014. DOI:10.1038/nrrheum.2014.47.

294 Petersen KK, Arendt-Nielsen L. Chronic Postoperative Pain After Joint Replacement. *Pain* 2016.

295 Wylde V, Hewlett S, Learmonth ID, Dieppe P. Persistent pain after joint replacement: Prevalence, sensory qualities, and postoperative determinants. *Pain* 2011. DOI:10.1016/j.pain.2010.11.023.

296 Sharma N. Atypical Analgesic in Chronic Pain Management. 2019; published online Aug 27. https://www.researchgate.net/publication/335421056_ATYPICAL_ANALGESIC_IN_CHRONIC_PAIN_MANAGEMENT.

297 Moore RA, Derry S, Aldington D, Cole P, Wiffen PJ. Amitriptyline for neuropathic pain in adults. *Cochrane Database Syst Rev* 2015; **7**: CD008242.

298 NICE. Neuropathic pain in adults : pharmacological management in non-specialist settings. *NICE Guidel* 2013. DOI:10.1016/S0735-1097(14)60199-6.

299 Caudill-Slosberg MA, Schwartz LM, Woloshin S. Office visits and analgesic prescriptions for musculoskeletal pain in US: 1980 vs. 2000. *Pain* 2004. DOI:10.1016/j.pain.2004.03.006.

300 Onghena P, Van Houdenhove B. Antidepressant-induced analgesia in chronic non-malignant pain: a meta-analysis of 39 placebo-controlled studies. *Pain* 1992. DOI:10.1016/0304-3959(92)90144-Z.

301 Grant Macfarlane J, Jalali S, Grace EM. Trimipramine in rheumatoid arthritis: A randomized double-blind trial in relieving pain and joint tenderness. *Curr Med Res Opin* 1986. DOI:10.1185/03007998609110424.

302 Chappell AS, Ossanna MJ, Liu-Seifert H, *et al.* Duloxetine, a centrally acting analgesic, in the treatment of patients with osteoarthritis knee pain: A 13-week, randomized, placebo-controlled trial. *Pain* 2009. DOI:10.1016/j.pain.2009.06.024.

303 Perrot S. Should we switch from analgesics to the concept of 'Pain Modifying Analgesic Drugs (PMADs)' in osteoarthritis and rheumatic chronic pain conditions? Pain. 2009. DOI:10.1016/j.pain.2009.07.037.

304 Pancrazio JJ, Kamatchi GL, Roscoe AK, Lynch C. Inhibition of neuronal Na+ channels by antidepressant drugs. *J Pharmacol Exp Ther* 1998.

305 Sarzi Puttini P, Cazzola M, Boccassini L, *et al.* A comparison of dothiepin versus placebo in the treatment of pain in rheumatoid arthritis and the association of pain with depression. *J Int Med Res* 1988. DOI:10.1177/030006058801600502.

306 Frank RG, Kashani JH, Parker JC, *et al.* Antidepressant analgesia in rheumatoid arthritis. *J Rheumatol* 1988.

307 Ash G, Dickens CM, Creed FH, Jayson MIV, Tomenson B. The effects of dothiepin on subjects with rheumatoid arthritis and depression. *Rheumatology* 1999. DOI:10.1093/rheumatology/38.10.959.

308 Maizels M, McCarberg B. Antidepressants and antiepileptic drugs for chronic non-cancer pain. Am. Fam. Physician. 2005. DOI:10.4102/safp.v48i3.534.

309 Gray P. Pregabalin in the management of central neuropathic pain. *Expert Opin Pharmacother* 2007. DOI:10.1517/14656566.8.17.3035.

310 O'Connor AB, Dworkin RH. Treatment of Neuropathic Pain: An Overview of Recent Guidelines. *Am J Med* 2009. DOI:10.1016/j.amjmed.2009.04.007.

311 Rahman W, Bauer CS, Bannister K, Vonsy JL, Dolphin AC, Dickenson AH. Descending serotonergic facilitation and the antinociceptive effects of pregabalin in a rat model of osteoarthritic pain. *Mol Pain* 2009. DOI:10.1186/1744-8069-5-45.

312 Clarke H, Pagé GM, McCartney CJL, *et al.* Pregabalin reduces postoperative opioid consumption and pain for 1 week after hospital discharge, but does not affect function at 6 weeks or 3 months after total hip arthroplasty. *Br J Anaesth* 2015. DOI:10.1093/bja/aev363.

313 Li F, Ma J, Kuang M, *et al.* The efficacy of pregabalin for the management of postoperative pain in primary total knee and hip arthroplasty: A meta-analysis. J. Orthop. Surg. Res. 2017. DOI:10.1186/s13018-017-0540-0.

314 Rose MA, Kam PCA. Gabapentin: Pharmacology and its use in pain management. Anaesthesia. 2002. DOI:10.1046/j.0003-2409.2001.02399.x.

315 Chouinard G, Beauclair L, Belanger MC. Gabapentin: Long-term antianxiety and hypnotic effects in psychiatric patients with comorbid anxiety-related disorders [1]. Can. J. Psychiatry. 1998.

316 Han C, Li X dan, Jiang H qiang, Ma J xiong, Ma X long. The use of gabapentin in the management of postoperative pain after total hip arthroplasty: A meta-analysis of randomised controlled trials. J. Orthop. Surg. Res. 2016. DOI:10.1186/s13018-016-0412-z.

317 Serpell MG. Gabapentin in neuropathic pain syndromes: A randomised, double-blind, placebo-controlled trial. *Pain* 2002. DOI:10.1016/S0304-3959(02)00255-5.

318 Klompas M, Cocoros NM, Menchaca JT, *et al.* State & local chronic disease

surveillance using electronic health record systems. *Am J Public Health* 2017. DOI:10.2105/AJPH.2017.303874.

319 Prentice HA, Chan PH, Namba RS, Inacio MC, Sedrakyan A, Paxton EW. Association of Type and Frequency of Postsurgery Care with Revision Surgery after Total Joint Replacement. *Perm J* 2019. DOI:10.7812/TPP/18.314.

320 Warren LR, Clarke J, Arora S, Darzi A. Improving data sharing between acute hospitals in England: An overview of health record system distribution and retrospective observational analysis of inter-hospital transitions of care. *BMJ Open* 2019. DOI:10.1136/bmjopen-2019-031637.

321 Benson T. Why general practitioners use computers and hospital doctors do not—Part 1: Incentives. *BMJ* 2002. DOI:10.1136/bmj.325.7372.1086.

322 McMillan B, Eastham R, Brown B, Fitton R, Dickinson D. Primary care patient records in the United Kingdom: Past, present, and future research priorities. J. Med. Internet Res. 2018. DOI:10.2196/11293.

323 Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat. Rev. Genet. 2011. DOI:10.1038/nrg2999.

324 Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: Towards better research applications and clinical care. Nat. Rev. Genet. 2012. DOI:10.1038/nrg3208.

325 Inacio MCS, Cafri G, Funahashi TT, Maletis GB, Paxton EW. Type and frequency of healthcare encounters can predict poor surgical outcomes in anterior cruciate ligament reconstruction patients. *Int J Med Inform* 2016.

DOI:10.1016/j.ijmedinf.2016.03.005.

326 Inacio MCS, Pratt NL, Roughead EE, Graves SE. Using Medications for Prediction of Revision after Total Joint Arthroplasty. *J Arthroplasty* 2015. DOI:10.1016/j.arth.2015.06.009.

327 van der Waal JM, Bot SD, Terwee CB, *et al.* The incidences of and consultation rate for lower extremity complaints in general practice. *Ann Rheum Dis* 2006; **65**: 809–15.

328 NHS England. NHS England » Musculoskeletal conditions. NHS Engl. 2020.

329 Hagen KB, Bjørndal A, Uhlig T, Kvien TK. A population study of factors associated with general practitioner consultation for non-inflammatory musculoskeletal pain. *Ann Rheum Dis* 2000. DOI:10.1136/ard.59.10.788.

330 Mitchell HL, Carr AJ, Scott DL. The management of knee pain in primary care: factors associated with consulting the GP and referrals to secondary care. *Rheumatol* 2006; **45**: 771–6.

331 Pinedo-Villanueva R, Turner D, Raftery JP, Cooper C, Arden NK. Primary care costs attributable to osteoarthritis before total hip replacement. *Osteoarthr Cartil* 2014. DOI:10.1016/j.joca.2014.02.404.

332 Paxton EW, Chan P, Love R, Kiley M-L, Namba RS, Inacio MCS. Automated EHR Surveillance System for Early Identification of Total Joint Replacement Failures. In: AAOS 2016 Annual Meeting. 2016. http://aaos2016.conferencespot.org/61513-aaos-1.2968416/t002-1.2975366/f002-1.2975367/a024-1.2976971/paper-191-1.2976981.

333 Sripa P, Hayhoe B, Majeed A, Greenfield G, Garg P. Impact of GP gatekeeping on quality of care, and health outcomes, use, and expenditure. Br. J. Gen. Pract. 2019. DOI:10.3399/bjgp19X702209.

334 Lin CT, Albertson G, Price D, Swaney R, Anderson S, Anderson RJ. Patient desire and reasons for specialist referral in a gatekeeper-model managed care plan. *Am J Manag Care* 2000.

335 McHugh GA, Campbell M, Luker KA. GP referral of patients with osteoarthritis for consideration of total joint replacement: A longitudinal study. *Br J Gen Pract* 2011. DOI:10.3399/bjgp11X588420.

336 Fitzpatrick R, Norquist JM, Reeves BC, Morris RW, Murray DW, Gregg PJ. Equity and need when waiting for total hip replacement surgery. J. Eval. Clin. Pract. 2004. DOI:10.1111/j.1365-2753.2003.00448.x.

337 Hudak PL, Grassau P, Glazier RH, *et al.* 'Not everyone who needs one is going to get one': The influence of medical brokering on patient candidacy for total joint arthroplasty. *Med Decis Mak* 2008. DOI:10.1177/0272989X08318468.

338 Birrell F, Afzal C, Nahit E, *et al.* Predictors of hip joint replacement in new attenders in primary care with hip pain. *Br J Gen Pract* 2003.

339 for Health NI, Excellence C. Osteoarthritis: care and management | Guidance and guidelines | NICE. *Nice Cg177* 2014.

340 Ballantyne PJ, Gignac MAM, Hawker GA. A patient-centered perspective on surgery avoidance for hip or knee arthritis: Lessons for the future. *Arthritis Care Res* 2007. DOI:10.1002/art.22472.

341 British Orthopaedic Association. Primary Total Hip Replacement: A guide to good practice. 2012. https://www.britishhipsociety.com/uploaded/Blue Book 2012 fsh nov 2012.pdf (accessed July 9, 2020).

342 NICE. Joint replacement (primary): hip, knee and shoulder. NICE Guidel. 2020.

343 Jørgensen CK, Olesen F. Predictors for referral to physiotherapy from general practice. *Scand J Prim Health Care* 2001. DOI:10.1080/028134301300034684.

344 Downie F, McRitchie C, Monteith W, Turner H. Physiotherapist as an alternative to a GP for musculoskeletal conditions: a 2-year service evaluation of UK primary care data. *Br J Gen Pract* 2019. DOI:10.3399/bjgp19x702245.

345 NHS. Recovery - Hip replacement. 2019. https://www.nhs.uk/conditions/hip-replacement/recovery/ (accessed July 9, 2020).

346 D. P-A, M. J. The effects of cement on all-cause mortality following total hip arthroplasty: The press-up cohort. *Osteoarthr Cartil* 2013.

347 Hughes RE, Batra A, Hallstrom BR. Arthroplasty registries around the world: valuable sources of hip implant revision risk data. Curr. Rev. Musculoskelet. Med. 2017. DOI:10.1007/s12178-017-9408-5.

348 Allepuz A, Serra-Sutton V, Martínez O, *et al.* Arthroplasty registers as post-marketing surveillance systems: The Catalan Arthroplasty Register. *Rev Española Cirugía Ortopédica y Traumatol (English Ed* 2013. DOI:10.1016/j.recote.2012.12.004.

349 Allepuz A, Martinez O, Tebe C, *et al.* Joint Registries as Continuous Surveillance Systems: The Experience of the Catalan Arthroplasty Register (RACat). *J Arthroplast* 2013. DOI:10.1016/j.arth.2013.07.048.

350 Arias-De La Torre J, Domingo L, Martínez O, *et al.* Evaluation of the effectiveness of hip and knee implant models used in Catalonia: A protocol for a prospective registry-based study. *J Orthop Surg Res* 2019. DOI:10.1186/s13018-019-1087-z.

351 Arias-de la Torre J, Pons-Cabrafiga M, Valderas J, *et al.* Influence of Hospital Volume of Procedures by Year on the Risk of Revision of Total Hip and Knee Arthroplasties: A Propensity Score-Matched Cohort Study. *J Clin Med* 2019. DOI:10.3390/jcm8050670.

352 SIDIAP. Population and Setting. https://www.sidiap.org/index.php/database/general-details (accessed Aug 29, 2020).

353 Ramos R, Ballo E, Marrugat J, *et al.* Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. *Rev Esp Cardiol (Engl Ed)* 2012; **65**: 29–37.

354 Garcia-Gil Mdel M, Hermosilla E, Prieto-Alhambra D, *et al.* Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). *Inf Prim Care* 2011; **19**: 135–45.

355 Perez-Saez MJ, Prieto-Alhambra D, Barrios C, *et al.* Increased hip fracture and mortality in chronic kidney disease individuals: The importance of competing risks. *Bone* 2014; **73C**: 154–9.

356 Prieto-Alhambra D, Premaor MO, Avilés FF, *et al.* Relationship between mortality and BMI after fracture: a population-based study of men and women aged ≥40 years. *J Bone Miner Res* 2014; **29**: 1737–44.

357 Reyes C, Estrada P, Nogues X, *et al.* The impact of common co-morbidities (as measured using the Charlson index) on hip fracture risk in elderly men: a population-based cohort study. *Osteoporos Int* 2014; **25**: 1751–8.

358 Reyes C, Garcia-Gil M, Elorza JM, *et al.* Socioeconomic status and its association with the risk of developing hip fractures: A region-wide ecological study. *Bone* 2014; **73C**: 127–31.

359 Wilson N, Sanchez-Riera L, Morros R, *et al.* Drug utilization in patients with OA: a population-based study. *Rheumatol* 2014. DOI:10.1093/rheumatology/keu403.

360 Martinez-Laguna D, Tebe C, Javaid MK, *et al.* Incident type 2 diabetes and hip fracture risk: a population-based matched cohort study. *Osteoporos Int* 2015; **26**: 827–33.

361 Prieto-Alhambra D, Judge A, Javaid MK, Cooper C, Diez-Perez A, Arden NK. Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. *Ann Rheum Dis* 2014; **73**: 1659–64.

362 Pages-Castella A, Carbonell-Abella C, Aviles FF, *et al.* 'Burden of osteoporotic fractures in primary health care in Catalonia (Spain): a population-based study'. *BMC Musculoskelet Disord* 2012; **13**: 79.

363 Fina-Aviles F, Medina-Peralta M, Mendez-Boo L, *et al.* The descriptive epidemiology of rheumatoid arthritis in Catalonia: a retrospective study using routinely collected data. *Clin Rheumatol* 2014. DOI:10.1007/s10067-014-2801-1.

364 Munoz-Ortego J, Vestergaard P, Rubio JB, *et al.* Ankylosing spondylitis is

associated with an increased risk of vertebral and nonvertebral clinical fractures: a population-based cohort study. *J Bone Min Res* 2014; **29**: 1770–6.

365 Ulrich SD, Seyler TM, Bennett D, *et al.* Total hip arthroplasties: what are the reasons for revision? *Int Orthop* 2008; **32**: 597–604.

366 Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes. *BMJ* 2010. DOI:10.1136/bmj.b5087.

367 Bessette BJ, Fassier F, Tanzer M, Brooks CE. Total hip arthroplasty in patients younger than 21 years: A minimum 10-year follow-up. *Can J Surg* 2003.

368 Jakobsen SS, Overgaard S, Søballe K, *et al.* The interface between periacetabular osteotomy, hip arthroscopy and total hip arthroplasty in the young adult hip. *EFORT Open Rev* 2018. DOI:10.1302/2058-5241.3.170042.

369 Kim YH, Oh SH, Kim JS, Koo KH. Contemporary total hip arthroplasty with and without cement in patients with osteonecrosis of the femoral head. *J Bone Jt Surg - Ser A* 2003. DOI:10.2106/00004623-200304000-00014.

370 Ong KL, Lau E, Suggs J, Kurtz SM, Manley MT. Risk of subsequent revision after primary and revision total joint arthroplasty. *Clin Orthop Relat Res* 2010. DOI:10.1007/s11999-010-1399-0.

371 Bayliss LE, Culliford D, Monk AP, *et al.* The effect of patient age at intervention on risk of implant revision after total replacement of the hip or knee: a population-based cohort study. *Lancet* 2017. DOI:10.1016/S0140-6736(17)30059-4.

372 Berry DJ, Harmsen WS, Cabanela ME, Morrey BF. Twenty-five-year

survivorship of two thousand consecutive primary Charnley total hip replacements: Factors affecting survivorship of acetabular and femoral components. *J Bone Jt Surg - Ser A* 2002. DOI:10.2106/00004623-200202000-00002.

373 de Kam DCJ, Gardeniers JWM, Veth RPH, Schreurs BW. Total hip arthroplasty in patients under 40 years of age and, if indicated, reconstruction of acetabular defects with bone impaction grafting. *Ned Tijdschr Geneeskd* 2010.

374 World Health Organization. The Consensus document. The Bone and Joint Decade 2000-2010. Inaugural Meeting 17 and 18 April 1998, Lund, Sweden. *Acta Orthop Scand* 1998; **69**: 67–86.

375 World Health Organization. ATC/DDD Index 2013. 2013. http://www.whocc.no/atc_ddd_index/.

376 St. Onge EL, Miller SA. Pain associated with diabetic peripheral neuropathy: A review of available treatments. P T. 2008.

377 Kringos DS, Boerma WGW, Hutchinson A, Saltman RD, Saltman RB. Building primary care in a changing Europe - Case studies. 2015 DOI:10.1002/we.193.

378 Corbacho B, Pinto-Prades JL. Health economic decision-making: A comparison between UK and Spain. Br. Med. Bull. 2012. DOI:10.1093/bmb/lds017.

379 Bordini B, Stea S, De Clerico M, Strazzari S, Sasdelli A, Toni A. Factors affecting aseptic loosening of 4750 total hip arthroplasties: multivariate survival analysis. *BMC Musculoskelet Disord* 2007; **8**: 69.

380 Skelly A, Dettori J, Brodt E. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J* 2012. DOI:10.1055/s-0031-1298595.

381 Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2003.

382 Roder C, Bach B, Berry DJ, Eggli S, Langenhahn R, Busato A. Obesity, age, sex, diagnosis, and fixation mode differently affect early cup failure in total hip arthroplasty: a matched case-control study of 4420 patients. *J Bone Jt Surg Am* 2010; **92**: 1954–63.

383 Santaguida PL, Hawker GA, Hudak PL, *et al.* Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. *Can J Surg* 2008; **51**: 428–36.

384 Roder C, Eggli S, Munger P, Melloh M, Busato A. Patient characteristics differently affect early cup and stem loosening in THA: a case-control study on 7,535 patients. *Int Orthop* 2008; **32**: 33–8.

385 Munger P, Roder C, Ackermann-Liebrich U, Busato A. Patient-related risk factors leading to aseptic stem loosening in total hip arthroplasty: a case-control study of 5,035 patients. *Acta Orthop* 2006; **77**: 567–74.

386 Poon PC, Rennie J, Gray DH. Review of total hip replacement. The Middlemore Hospital experience, 1980-1991. N. Z. Med. J. 2001.

387 Johnsen SP, Sorensen HT, Lucht U, Soballe K, Overgaard S, Pedersen AB. Patient-related predictors of implant failure after primary total hip replacement in the initial, short- and long-terms. A nationwide Danish follow-up study including 36,984 patients. *J Bone Jt Surg Br* 2006; **88**: 1303–8.

388 Young NL, Cheah D, Waddell JP, Wright JG. Patient characteristics that affect the outcome of total hip arthroplasty: a review. *Can J Surg* 1998; **41**: 188–95.

389 Federman AD, Litke A, Morrison RS. Association of age with analgesic use for back and joint disorders in outpatient settings. *Am J Geriatr Pharmacother* 2006. DOI:10.1016/j.amjopharm.2006.12.009.

390 Rhon DI, Snodgrass SJ, Cleland JA, Sissel CD, Cook CE. Predictors of chronic prescription opioid use after orthopedic surgery: derivation of a clinical prediction rule. *Perioper Med* 2018. DOI:10.1186/s13741-018-0105-8.

391 Rajamäki TJ, Puolakka PA, Hietaharju A, Moilanen T, Jämsen E. Predictors of the use of analgesic drugs 1 year after joint replacement: a single-center analysis of 13,000 hip and knee replacements. *Arthritis Res Ther* 2020. DOI:10.1186/s13075-020-02184-1.

392 Clarke H, Soneji N, Ko DT, Yun L, Wijeysundera DN. Rates and risk factors for prolonged opioid use after major surgery: Population based cohort study. *BMJ* 2014. DOI:10.1136/bmj.g1251.

393 Ip HYV, Abrishami A, Peng PWH, Wong J, Chung F. Predictors of Postoperative Pain and Analgesic Consumption. *Anesthesiology* 2009. DOI:10.1097/aln.0b013e3181aae87a.

394 Armijo JA, Pena MA, Adín J, Vega-Gil N. Association between patient age and gabapentin serum concentration-to-dose ratio: A preliminary multivariate analysis. *Ther Drug Monit* 2004. DOI:10.1097/00007691-200412000-00008.

395 Schmader KE, Baron R, Haanpää ML, *et al.* Treatment considerations for elderly and frail patients with neuropathic pain. Mayo Clin. Proc. 2010. DOI:10.4065/mcp.2009.0646.

396 Torrance N, Smith BH, Bennett MI, Lee AJ. The Epidemiology of Chronic Pain of Predominantly Neuropathic Origin. Results From a General Population Survey. *J Pain* 2006. DOI:10.1016/j.jpain.2005.11.008.

397 Fick DM, Semla TP, Steinman M, *et al.* American Geriatrics Society 2019 Updated AGS Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults. *J Am Geriatr Soc* 2019. DOI:10.1111/jgs.15767.

398 Wammes JJG, Van Der Wees PJ, Tanke MAC, Westert GP, Jeurissen PPT. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open. 2018. DOI:10.1136/bmjopen-2018-023113.

399 Hu T, Dattani ND, Cox KA, *et al.* Effect of comorbidities and medications on frequency of primary care visits among older patients. *Can Fam Physician* 2017.

400 Rosella LC, Kornas K, Yao Z, *et al.* Predicting high health care resource utilization in a single-payer public health care system: Development and validation of the high resource user population risk tool. *Med Care* 2018. DOI:10.1097/MLR.0000000000000837.

401 van den Bogaart EHA, Spreeuwenberg MD, Kroese MEAL, van den Boogaart MW, Boymans TAEJ, Ruwaard D. Referral decisions and its predictors related to orthopaedic care. A retrospective study in a novel primary care setting. *PLoS One* 2020. DOI:10.1371/journal.pone.0227863.

402 McBride D, Hardoon S, Walters K, Gilmour S, Raine R. Explaining variation in referral from primary to secondary care: Cohort study. *BMJ* 2010. DOI:10.1136/bmj.c6267.

403 Walters K, Iliffe S, Orrell M. An exploration of help-seeking behaviour in older people with unmet needs. *Fam Pract* 2001. DOI:10.1093/fampra/18.3.277.

404 Jameson SS, Baker PN, Mason J, Porter ML, Deehan DJ, Reed MR. Independent predictors of revision following metal-on-metal hip resurfacing: a retrospective cohort study using National Joint Registry data. *J Bone Jt Surg Br* 2012; **94**: 746–54.

405 Campbell CI, Weisner C, LeResche L, *et al.* Age and gender trends in long-term opioid analgesic use for noncancer pain. *Am J Public Health* 2010. DOI:10.2105/AJPH.2009.180646.

406 Wise EA, Price DD, Myers CD, Heft MW, Robinson ME. Gender role expectations of pain: Relationship to experimental pain perception. *Pain* 2002. DOI:10.1016/S0304-3959(01)00473-0.

407 Robinson ME, Dannecker EA, George SZ, Otis J, Atchison JW, Fillingim RB. Sex differences in the associations among psychological factors and pain report: A novel psychophysical study of patients with chronic low back pain. *J Pain* 2005. DOI:10.1016/j.jpain.2005.02.007.

408 Wandner LD, Scipio CD, Hirsh AT, Torres CA, Robinson ME. The perception of pain in others: How gender, race, and age influence pain expectations. *J Pain* 2012. DOI:10.1016/j.jpain.2011.10.014.

409 Fernandes GS, Valdes AM, Walsh DA, Zhang W, Doherty M. Neuropathic-like knee pain and associated risk factors: a cross-sectional study in a UK community sample. *Arthritis Res Ther* 2018. DOI:10.1186/s13075-018-1717-6.

410 Johannessen Landmark C, Beiske G, Baftiu A, Burns ML, Johannessen SI. Experience from therapeutic drug monitoring and gender aspects of gabapentin and pregabalin in clinical practice. *Seizure* 2015. DOI:10.1016/j.seizure.2015.02.017.

411 Anderson GD. Chapter 1 Gender Differences in Pharmacological Response. Int. Rev. Neurobiol. 2008. DOI:10.1016/S0074-7742(08)00001-9.

412 Whitley HP, Lindsey W. Sex-based differences in drug activity. Am. Fam. Physician. 2009.

413 Karlson EW, Daltroy LH, Liang MH, Eaton HE, Katz JN. Gender differences in patient preferences may underlie differential utilization of elective surgery. *Am J Med* 1997. DOI:10.1016/S0002-9343(97)00050-8.

414 Mota REM, Tarricone R, Ciani O, Bridges JFP, Drummond M. Determinants of demand for total hip and knee arthroplasty: A systematic literature review. *BMC Health Serv Res* 2012. DOI:10.1186/1472-6963-12-225.

415 Uhlig T, Hagen KB, Kvien TK. Why do patients with chronic musculoskeletal disorders consult their primary care physicians? Curr. Opin. Rheumatol. 2002. DOI:10.1097/00002281-200203000-00005.

416 MacKay C, Canizares M, Davis AM, Badley EM. Health care utilization for musculoskeletal disorders. *Arthritis Care Res* 2010. DOI:10.1002/acr.20064.

417 Korten AE, Jacomb PA, Jiao Z, *et al.* Predictors of GP service use: A community survey of an elderly Australian sample. *Aust N Z J Public Health* 1998. DOI:10.1111/j.1467-842X.1998.tb01447.x.

418 Akpala CO, Curran AP, Simpson J. Physiotherapy in general practice: Patterns of utilisation. *Public Health* 1988. DOI:10.1016/S0033-3506(88)80069-6.

419 Inoue K, Ushiyama T, Tani Y, Hukuda S. Sociodemographic factors and failure of hip arthroplasty. *Int Orthop* 1999; **23**: 330–3.

420 Clement ND, Muzammil A, Macdonald D, Howie CR, Biant LC. Socioeconomic status affects the early outcome of total hip replacement. *J Bone Joint Surg Br* 2011; **93**: 464–9.

421 McBeth J, Jones K. Epidemiology of chronic musculoskeletal pain. Best Pract. Res. Clin. Rheumatol. 2007. DOI:10.1016/j.berh.2007.03.003.

422 Bonathan C, Hearn L, Williams AC de C. Socioeconomic status and the course and consequences of chronic pain. *Pain Manag* 2013. DOI:10.2217/pmt.13.18.

423 Lungu E, Maftoon S, Vendittoli PA, Desmeules F. A systematic review of preoperative determinants of patient-reported pain and physical function up to 2 years following primary unilateral total hip arthroplasty. Orthop. Traumatol. Surg. Res. 2016. DOI:10.1016/j.otsr.2015.12.025.

424 Wakefield JRH, Sani F, Madhok V, Norbury M, Dugard P. The pain of low status: The relationship between subjective socio-economic status and analgesic prescriptions in a Scottish community sample. *Psychol Heal Med* 2016. DOI:10.1080/13548506.2015.1009377.

425 Prieto-Alhambra D, Judge A, Javaid M, *et al.* Predictors of poor outcome (high use of non-steroidal anti-inflammatory drugs) at year 1 following total knee/hip arthroplasty: the press-up cohort study. *Osteoarthr Cartil* 2013; **21**: S155.

426 VanDenKerkhof EG, Mann EG, Torrance N, Smith BH, Johnson A, Gilron I. An epidemiological study of neuropathic pain Symptoms in Canadian adults. *Pain Res Manag* 2016. DOI:10.1155/2016/9815750.

427 Martikainen P, Mäkelä P, Koskinen S, Valkonen T. Income differences in mortality: A register-based follow-up study of three million men and women. *Int J Epidemiol* 2001. DOI:10.1093/ije/30.6.1397.

428 Brekke M, Hjortdahl P, Kvien TK. Severity of musculoskeletal pain: Relations to socioeconomic inequality. *Soc Sci Med* 2002. DOI:10.1016/S0277-9536(01)00018-1.

429 Regenstein M, Lantz PM. Socioeconomic Status and Health Care. In: International Encyclopedia of the Social & Behavioral Sciences: Second Edition. 2015. DOI:10.1016/B978-0-08-097086-8.14044-9.

430 Åhs A, Burell G, Westerling R. Care or not care-that is the question: Predictors of healthcare utilisation in relation to employment status. *Int J Behav Med* 2012. DOI:10.1007/s12529-010-9129-2.

431 Fitzpatrick T, Rosella LC, Calzavara A, *et al.* Looking beyond Income and Education: Socioeconomic Status Gradients among Future High-Cost Users of Health Care. *Am J Prev Med* 2015. DOI:10.1016/j.amepre.2015.02.018.

432 Lemstra M, Mackenbach J, Neudorf C, Nannapaneni U. High health care utilization and costs associated with lower socio-economic status: Results from a linked dataset. *Can J Public Heal* 2009. DOI:10.1007/bf03405536.

433 Bozic KJ, Lau E, Ong K, *et al.* Risk factors for early revision after primary total hip arthroplasty in Medicare patients. *Clin Orthop Relat Res* 2014; **472**: 449–54.

434 Jamsen E, Peltola M, Eskelinen A, Lehto MU. Comorbid diseases as predictors of survival of primary total hip and knee replacements: a nationwide register-based study of 96 754 operations on patients with primary osteoarthritis. *Ann Rheum Dis* 2013; **72**: 1975–82.

435 Pedersen AB, Mehnert F, Johnsen SP, Sorensen HT. Risk of revision of a total hip replacement in patients with diabetes mellitus: a population-based follow up study. *J Bone Jt Surg Br* 2010; **92**: 929–34.

436 Gordon M, Stark A, Skoldenberg OG, Karrholm J, Garellick G. The influence of comorbidity scores on re-operations following primary total hip replacement: comparison and validation of three comorbidity measures. *Bone Jt J* 2013; **95-B**: 1184–91.

437 Armitage JN, Van Der Meulen JH. Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score. *Br J Surg* 2010. DOI:10.1002/bjs.6930.

438 Antonelli B, Chen AF. Reducing the risk of infection after total joint arthroplasty: preoperative optimization. *Arthroplasty* 2019. DOI:10.1186/s42836-019-0003-7.

439 Pulido L, Ghanem E, Joshi A, Purtill JJ, Parvizi J. Periprosthetic joint infection: The incidence, timing, and predisposing factors. *Clin Orthop Relat Res* 2008. DOI:10.1007/s11999-008-0209-4.

440 Korol E, Johnston K, Waser N, *et al.* A systematic review of risk factors associated with surgical site infections among surgical patients. PLoS One. 2013. DOI:10.1371/journal.pone.0083743.

441 Aggarwal VK, Tischler EH, Post ZD, Kane I, Orozco FR, Ong A. Patients with atrial fibrillation undergoing total joint arthroplasty increase hospital burden. *J Bone Jt Surg - Ser A* 2013. DOI:10.2106/JBJS.L.00882.

442 Davis JA, Robinson RL, Le TK, Xie J. Incidence and impact of pain conditions and comorbid illnesses. *J Pain Res* 2011. DOI:10.2147/JPR.S24170.

443 Singh JA, Lewallen DG. Medical comorbidity is associated with persistent index hip pain after total hip arthroplasty. *Pain Med (United States)* 2013. DOI:10.1111/pme.12153.

444 Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. *J Clin Epidemiol* 2008. DOI:10.1016/j.jclinepi.2008.01.006.

445 Schmolders J, Friedrich MJ, Michel R, *et al.* Validation of the Charlson comorbidity index in patients undergoing revision total hip arthroplasty. *Int Orthop* 2015. DOI:10.1007/s00264-015-2810-y.

446 Huntley AL, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of multimorbidity and morbidity burden for use in primary care and community settings: A systematic review and guide. Ann. Fam. Med. 2012. DOI:10.1370/afm.1363.

447 Hopman P, Heins MJ, Rijken M, Schellevis FG. Health care utilization of patients with multiple chronic diseases in the Netherlands: Differences and underlying factors. *Eur J Intern Med* 2015. DOI:10.1016/j.ejim.2015.02.006.

448 Starfield B, Lemke KW, Bernhardt T, Foldes SS, Forrest CB, Weiner JP. Comorbidity: implications for the importance of primary care in 'case' management. *Ann Fam Med* 2003. DOI:10.1370/afm.1.

449 Barten DJJA, Smink A, Swinkels ICS, *et al.* Factors Associated With Referral to Secondary Care in Patients With Osteoarthritis of the Hip or Knee After Implementation of a Stepped-Care Strategy. *Arthritis Care Res* 2017. DOI:10.1002/acr.22935.

450 Marcum ZA, Perera S, Donohue JM, *et al.* Analgesic use for knee and hip osteoarthritis in community-dwelling elders. *Pain Med* 2011. DOI:10.1111/j.1526-4637.2011.01249.x.

451 Krijthe BP, Heeringa J, Hofman A, Franco OH, Stricker BH. Non-steroidal anti-inflammatory drugs and the risk of atrial fibrillation: A population-based follow-up study. *BMJ Open* 2014. DOI:10.1136/bmjopen-2013-004059.

452 Schmidt M, Christiansen CF, Mehnert F, Rothman KJ, Sørensen HT. Non-steroidal anti-inflammatory drug use and risk of atrial fibrillation or flutter: Population based case-control study. *BMJ* 2011. DOI:10.1136/bmj.d3450.

453 Braeken AM, Lochhaas-Gerlach JA, Gollish JD, Myles JD, Mackenzie TA. Determinants of 6-12 month postoperative functional status and pain after elective total hip replacement. *Int J Qual Heal Care* 1997. DOI:10.1093/intqhc/9.6.413.

454 Schjerning A-M, McGettigan P, Gislason G. Cardiovascular effects and safety of (non-aspirin) NSAIDs. *Nat Rev Cardiol* 2020. DOI:10.1038/s41569-020-0366-z.

455 Ortiz de Landaluce L, Carbonell P, Asensio C, Escoda N, López P, Laporte JR. Gabapentin and Pregabalin and Risk of Atrial Fibrillation in the Elderly: A Population-Based Cohort Study in an Electronic Prescription Database. *Drug Saf* 2018. DOI:10.1007/s40264-018-0695-6.

456 British National Formulary (BNF). British National Formulary: Pregablin. 2020. https://bnf.nice.org.uk/drug/pregabalin.html#cautions (accessed June 9, 2020).

457 British National Formulary (BNF). British National Formulary: Amitritptyline. 2020. https://bnf.nice.org.uk/drug/amitriptyline-hydrochloride.html#contraIndications (accessed June 9, 2020).

458 Dworkin RH, O'Connor AB, Audette J, *et al.* Recommendations for the pharmacological management of neuropathic pain: An overview and literature update. Mayo Clin. Proc. 2010. DOI:10.4065/mcp.2009.0649.

459 Skarshaug LJ, Kaspersen SL, Bjørngaard JH, Pape K. Changes in General Practitioners' consultation frequency over time for patients with hypertension or anxiety/depression symptoms: a 10-year follow-up of the Norwegian HUNT study. *Fam Pract* 2020. DOI:10.1093/fampra/cmz070.

460 NICE. Hypertension: Clinical management of primary hypertension in adults (NICE clinical guideline 127). *NICE* 2011.

461 Kassianos G, Arden C, Hogan S, Dew R, Fuat A. Current management of atrial fibrillation: An observational study in NHS primary care. *BMJ Open* 2013. DOI:10.1136/bmjopen-2013-003004.

462 Rajamäki TJ, Jämsen E, Puolakka PA, Nevalainen PI, Moilanen T. Diabetes is

associated with persistent pain after hip and knee replacement. Acta Orthop. 2015. DOI:10.3109/17453674.2015.1044389.

463 Jones CA, Cox V, Jhangri GS, Suarez-Almazor ME. Delineating the impact of obesity and its relationship on recovery after total joint arthroplasties. *Osteoarthr Cartil* 2012. DOI:10.1016/j.joca.2012.02.637.

464 May M, Schindler C. Clinically and pharmacologically relevant interactions of antidiabetic drugs. Ther. Adv. Endocrinol. Metab. 2016. DOI:10.1177/2042018816638050.

465 Dhatariya K, Levy N, Flanagan D, *et al.* Management of adults with diabetes undergoing surgery and elective procedures: Improving standards. 2016.

466 Meding JB, Klay M, Healy A, Ritter MA, Keating EM, Berend ME. The Prescreening History and Physical in Elective Total Joint Arthroplasty. *J Arthroplasty* 2007. DOI:10.1016/j.arth.2007.03.035.

467 Van Hecke O, Austin SK, Khan RA, Smith BH, Torrance N. Neuropathic pain in the general population: A systematic review of epidemiological studies. Pain. 2014. DOI:10.1016/j.pain.2013.11.013.

468 Barnard K, Peveler RC, Holt RIG. Antidepressant medication as a risk factor for type 2 diabetes and impaired glucose regulation. Diabetes Care. 2013. DOI:10.2337/dc13-0560.

469 British National Formulary (BNF). British National Formulary: Gabapentin. 2020. https://bnf.nice.org.uk/drug/gabapentin.html#cautions (accessed June 9, 2020).

470 NICE NG28. Type 2 diabetes in adults: management. *Natl Inst Heal Care Excell* 2015.

471 Wukich DK. Diabetes and its negative impact on outcomes in orthopaedic surgery. World J. Orthop. 2015. DOI:10.5312/wjo.v6.i3.331.

472 Singh JA, Lewallen D. Age, gender, obesity, and depression are associated with patient-related pain and function outcome after revision total hip arthroplasty. *Clin Rheumatol* 2009. DOI:10.1007/s10067-009-1267-z.

473 Blom AW, Artz N, Beswick AD, *et al.* Improving patients' experience and outcome of total joint replacement: the RESTORE programme. *Program Grants Appl Res* 2016. DOI:10.3310/pgfar04120.

474 Berger A, Bozic K, Stacey B, Edelsberg J, Sadosky A, Oster G. Patterns of pharmacotherapy and health care utilization and costs prior to total hip or total knee replacement in patients with osteoarthritis. *Arthritis Rheum* 2011. DOI:10.1002/art.30417.

475 Boogaard S, Heymans MW, Patijn J, *et al.* Predictors for persistent neuropathic pain - a delphi survey. *Pain Physician* 2011.

476 Bogoch ER, Olschewski E, Zangger P, Henke ML, Smythe HA. Increased tender point counts before and after total hip arthroplasty are associated with poorer outcomes but are not individually predictive. *J Arthroplasty* 2010. DOI:10.1016/j.arth.2009.08.005.

477 Duivenvoorden T, Vissers MM, Verhaar JAN, *et al.* Anxiety and depressive symptoms before and after total hip and knee arthroplasty: A prospective

multicentre study. *Osteoarthr Cartil* 2013. DOI:10.1016/j.joca.2013.08.022.

478 Keeley P, Creed F, Tomenson B, Todd C, Borglin G, Dickens C. Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain. *Pain* 2008. DOI:10.1016/j.pain.2007.05.015.

479 Pini S, Piccinelli M, Zimmermann-Tansella C. Social problems as factors affecting medical consultation: a comparison between general practice attenders and community probands with emotional distress. *Psychol Med* 1995.

480 Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: A systematic review. Man. Ther. 2010. DOI:10.1016/j.math.2009.12.004.

481 Kolundžić R, Šulentić M, Smerdelj M, Orlić D, Trkulja V. Stability of endler cementless polyethylene acetabular cup: Long-term follow-up. *Croat Med J* 2005.

482 Jameson SS, Baker PN, Mason J, *et al.* Independent predictors of failure up to 7.5 years after 35 386 single-brand cementless total hip replacements: a retrospective cohort study using National Joint Registry data. *Bone Jt J* 2013; **95-B**: 747–57.

483 Wright EA, Katz JN, Baron JA, *et al.* Risk factors for revision of primary total hip replacement: Results from a national case-control study. *Arthritis Care Res (Hoboken)* 2012; **64**: 1879–85.

484 Sayed-Noor AS, Mukka S, Mohaddes M, Kärrholm J, Rolfson O. Body mass index is associated with risk of reoperation and revision after primary total hip arthroplasty: a study of the Swedish Hip Arthroplasty Register including 83,146 patients. *Acta Orthop* 2019. DOI:10.1080/17453674.2019.1594015.

485 Mouchti S, Whitehouse MR, Sayers A, Hunt LP, Macgregor A, Blom AW. The Association of Body Mass Index with Risk of Long-Term Revision and 90-Day Mortality Following Primary Total Hip Replacement: Findings from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. *J Bone Jt Surg - Am Vol* 2018. DOI:10.2106/JBJS.18.00120.

486 Marras F, Leali PT. The role of drugs in bone pain. Clin. Cases Miner. Bone Metab. 2016. DOI:10.11138/ccmbm/2016.13.2.093.

487 Edwards HL, Mulvey MR, Bennett MI. Cancer-related neuropathic pain. Cancers (Basel). 2019. DOI:10.3390/cancers11030373.

488 Fallon MT, Colvin L. Neuropathic pain in cancer. *Br J Anaesth* 2013. DOI:10.1093/bja/aet208.

489 Mishra S, Bhatnagar S, Goyal GN, Rana SPS, Upadhya SP. A Comparative Efficacy of Amitriptyline, Gabapentin, and Pregabalin in Neuropathic Cancer Pain: A Prospective Randomized Double-Blind Placebo-Controlled Study. *Am J Hosp Palliat Med* 2012. DOI:10.1177/1049909111412539.

490 Mercadante S, Arcuri E, Tirelli W, Villari P, Casuccio A. Amitriptyline in neuropathic cancer pain in patients on morphine therapy: A randomized placebo-controlled, double-blind crossover study. *Tumori* 2002. DOI:10.1177/030089160208800310.

491 Dybvik E, Fosså SD, Furnes O, Lie SA, Trovik C. Long term risk for receiving a total hip replacement in cancer patients - A linkage study between the cancer register of Norway and the Norwegian arthroplasty register. *Orthop Proc* 2018; **Vol. 92-B,**. https://online.boneandjoint.org.uk/doi/abs/10.1302/0301-

620X.92BSUPP_IV.0920601b.

492 Chen DQ, Montgomery SR, Cancienne JM, Werner BC. Postoperative Venous Thromboembolism and other Complications after Anatomic Total Shoulder Arthroplasty in Patients With a History of Prostate or Breast Cancer. *J Am Acad Orthop Surg* 2020. DOI:10.5435/JAAOS-D-18-00777.

493 Rosas S, Tipton S, Luo TD, Plate JF, Willey JS, Emory CL. Complications and Costs Are Not Increased After Total Hip Arthroplasty in Patients With a History of Prostate Cancer. *J Arthroplasty* 2019. DOI:10.1016/j.arth.2019.06.052.

494 Clohisy JC, Harris WH. Matched-pair analysis of cemented and cementless acetabular reconstruction in primary total hip arthroplasty. *J Arthroplasty* 2001. DOI:10.1054/arth.2001.24447.

495 Crofford LJ. Use of NSAIDs in treating patients with arthritis. Arthritis Res. Ther. 2013. DOI:10.1186/ar4174.

496 Blackburn WD. Management of osteoarthritis and rheumatoid arthritis: Prospects and possibilities. In: American Journal of Medicine. 1996. DOI:10.1016/s0002-9343(97)89543-5.

497 van den Driest JJ, Bierma-Zeinstra SMA, Bindels PJ, Schiphof D. Amitriptyline for musculoskeletal complaints: A systematic review. Fam. Pract. 2017. DOI:10.1093/fampra/cmw134.

498 Lee YC, Nassikas NJ, Clauw DJ. The role of the central nervous system in the generation and maintenance of chronic pain in rheumatoid arthritis, osteoarthritis and fibromyalgia. Arthritis Res. Ther. 2011. DOI:10.1186/ar3306.

499 Emery P, Breedveld FC, Dougados M, Kalden JR, Schiff MH, Smolen JS. Early referral recommendation for newly diagnosed rheumatoid arthritis: Evidence based development of a clinical guide. *Ann Rheum Dis* 2002. DOI:10.1136/ard.61.4.290.

500 Almoallim H, Janoudi N, Attar SM, *et al.* Determining early referral criteria for patients with suspected inflammatory arthritis presenting to primary care physicians: A cross-sectional study. *Open Access Rheumatol Res Rev* 2017. DOI:10.2147/OARRR.S134780.

501 Clement ND, Breusch SJ, Biant LC. Lower limb joint replacement in rheumatoid arthritis. J. Orthop. Surg. Res. 2012. DOI:10.1186/1749-799X-7-27.

502 Visuri T, Turula KB, Pulkkinen P, Nevalainen J. Survivorship of hip prosthesis in primary arthrosis: Influence of bilaterality and interoperative time in 45,000 hip prostheses from the Finnish Endoprosthesis Register. *Acta Orthop Scand* 2002. DOI:10.1080/000164702320155266.

503 Jones CA, Voaklander DC, Johnston WC, Suarez-Almazor ME. The effect of age on pain, function, and quality of life after total hip and knee arthroplasty. *Arch Intern Med* 2001. DOI:10.1001/archinte.161.3.454.

504 Quintana JM, Escobar A, Aguirre U, Lafuente I, Arenaza JC. Predictors of health-related quality-of-life change after total hip arthroplasty. *Clin Orthop Relat Res* 2009. DOI:10.1007/s11999-009-0868-9.

505 Wong K, Phelan R, Kalso E, *et al.* Antidepressant drugs for prevention of acute and chronic postsurgical pain early evidence and recommended future directions. *Anesthesiology* 2014. DOI:10.1097/ALN.0000000000000307.

506 Hasija R, Kelly JJ, Shah N V., *et al.* Nerve injuries associated with total hip arthroplasty. J. Clin. Orthop. Trauma. 2018. DOI:10.1016/j.jcot.2017.10.011.

507 Borkhoff CM, Hawker GA, Wright JG. Patient gender affects the referral and recommendation for total joint arthroplasty. In: Clinical Orthopaedics and Related Research. 2011. DOI:10.1007/s11999-011-1879-x.

508 Waugh EJ, Badley EM, Borkhoff CM, *et al.* Primary care physicians' perceptions about and confidence in deciding which patients to refer for total joint arthroplasty of the hip and knee. *Osteoarthr Cartil* 2016. DOI:10.1016/j.joca.2015.09.017.

509 Hawker GA, Badley EM, Borkhoff CM, *et al.* Which patients are most likely to benefit from total joint arthroplasty? *Arthritis Rheum* 2013. DOI:10.1002/art.37901.

510 Furnes O, Lie SA, Espehaug B, Vollset SE, Engesaeter LB, Havelin LI. Hip disease and the prognosis of total hip replacements. A review of 53,698 primary total hip replacements reported to the Norwegian Arthroplasty Register 1987-99. *J Bone Jt Surg Br* 2001; **83**: 579–86.

511 Miller M, Stürmer T, Azrael D, Levin R, Solomon DH. Opioid analgesics and the risk of fractures in older adults with arthritis. *J Am Geriatr Soc* 2011. DOI:10.1111/j.1532-5415.2011.03318.x.

512 Rbia N, van der Vlies CH, Cleffken BI, Selles RW, Hovius SER, Nijhuis THJ. High Prevalence of Chronic Pain With Neuropathic Characteristics After Open Reduction and Internal Fixation of Ankle Fractures. *Foot Ankle Int* 2017. DOI:10.1177/1071100717712432.

513 Friesgaard KD, Gromov K, Knudsen LF, Brix M, Troelsen A, Nikolajsen L. Persistent pain is common 1 year after ankle and wrist fracture surgery: A register-based questionnaire study. *Br J Anaesth* 2016. DOI:10.1093/bja/aew069.

514 Jetté N, Lix LM, Metge CJ, Prior HJ, McChesney J, Leslie WD. Association of antiepileptic drugs with nontraumatic fractures: A population-based analysis. *Arch Neurol* 2011. DOI:10.1001/archneurol.2010.341.

515 Compston JE, McClung MR, Leslie WD. Osteoporosis. Lancet. 2019. DOI:10.1016/S0140-6736(18)32112-3.

516 Lindberg-Larsen M, Jørgensen CC, Solgaard S, Kjersgaard AG, Kehlet H. Increased risk of intraoperative and early postoperative periprosthetic femoral fracture with uncemented stems. *Acta Orthop* 2017. DOI:10.1080/17453674.2017.1302908.

517 Meldrum RD, Wurtz LD, Feinberg JR, Capello WN. Does smoking affect implant survivorship in total hip arthroplasty? A preliminary retrospective case series. *Iowa Orthop J* 2005.

518 Tischler EH, Ko LM, Chen AF, Maltenfort MG, Schroeder J, Austin MS. Smoking increases the rate of reoperation for infection within 90 days after primary total joint arthroplasty. J. Bone Jt. Surg. - Am. Vol. 2017. DOI:10.2106/JBJS.16.00311.

519 Matharu GS, Mouchti S, Twigg S, *et al.* The effect of smoking on outcomes following primary total hip and knee arthroplasty: a population-based cohort study of 117,024 patients. *Acta Orthop* 2019. DOI:10.1080/17453674.2019.1649510.

520 Palmer KT, Syddall H, Cooper C, Coggon D. Smoking and musculoskeletal

disorders: Findings from a British national survey. *Ann Rheum Dis* 2003. DOI:10.1136/ard.62.1.33.

521 Çelik SB, Can H, Sözmen MK, *et al.* Evaluation of the neuropathic pain in the smokers. *Agri* 2017. DOI:10.5505/agri.2017.68815.

522 Sood A, Ebbert JO, Schroeder DR, *et al.* Gabapentin for smoking cessation: A preliminary investigation of efficacy. *Nicotine Tob Res* 2007. DOI:10.1080/14622200601080307.

523 Herman AI, Waters AJ, McKee SA, Sofuoglu M. Effects of pregabalin on smoking behavior, withdrawal symptoms, and cognitive performance in smokers. *Psychopharmacology (Berl)* 2012. DOI:10.1007/s00213-011-2507-x.

524 Rosella LC, Fitzpatrick T, Wodchis WP, Calzavara A, Manson H, Goel V. High-cost health care users in Ontario, Canada: Demographic, socio-economic, and health status characteristics. *BMC Health Serv Res* 2014. DOI:10.1186/s12913-014-0532-2.

525 Møller AM, Pedersen T, Villebro N, Munksgaard A. Effect of smoking on early complications after elective orthopaedic surgery. *J Bone Jt Surg - Ser B* 2003. DOI:10.1302/0301-620X.85B2.13717.

526 Williams G, Daly M, Proude EM, *et al.* The influence of alcohol and tobacco use in orthopaedic inpatients on complications of surgery. *Drug Alcohol Rev* 2008. DOI:10.1080/09595230701711108.

527 Best MJ, Buller LT, Gosthe RG, Klika AK, Barsoum WK. Alcohol Misuse is an Independent Risk Factor for Poorer Postoperative Outcomes Following Primary

Total Hip and Total Knee Arthroplasty. *J Arthroplasty* 2015. DOI:10.1016/j.arth.2015.02.028.

528 Murphy L, Ng KWK, Su VCH, *et al.* Approach to the pharmacological management of chronic pain in patients with an alcohol use disorder. *J Pain Res* 2015. DOI:10.2147/JPR.S88900.

529 Witkiewitz K, Vowles KE. Alcohol and Opioid Use, Co-Use, and Chronic Pain in the Context of the Opioid Epidemic: A Critical Review. Alcohol. Clin. Exp. Res. 2018. DOI:10.1111/acer.13594.

530 Kawanabe K, Tanaka K, Tamura J, *et al.* Effect of alumina femoral head on clinical results in cemented total hip arthroplasty: Old versus current alumina. In: Journal of Orthopaedic Science. 2005. DOI:10.1007/s00776-005-0911-y.

531 Laupacis A, Bourne R, Rorabeck C, Feeny D, Tugwell P, Wong C. Comparison of total hip arthroplasty performed with and without cement. A randomized trial. *J Bone Jt Surg - Ser A* 2002. DOI:10.2106/00004623-200210000-00013.

532 Hooper GJ, Rothwell AG, Stringer M, Frampton C. Revision following cemented and uncemented primary total hip replacement: a seven-year analysis from the New Zealand Joint Registry. *J Bone Jt Surg Br* 2009; **91**: 451–8.

533 Ogino D, Kawaji H, Konttinen L, *et al.* Total hip replacement in patients eighty years of age and older. *J Bone Jt Surg - Ser A* 2008. DOI:10.2106/JBJS.G.00147.

534 Katz JN, Losina E, Barrett JA, *et al.* Association Between Hospital and Surgeon Procedure Volume and Outcomes of Total Hip Replacement in the United States Medicare Population. *J Bone Jt Surg* 2003; **85**: 27–32.

535 Manley M, Ong K, Lau E, Kurtz SM. Effect of volume on total hip arthroplasty revision rates in the United States Medicare population. *J Bone Jt Surg Am* 2008; **90**: 2446–51.

536 Losina E, Barrett J, Mahomed NN, Baron JA, Katz JN. Early failures of total hip replacement: effect of surgeon volume. *Arthritis Rheum* 2004; **50**: 1338–43.

537 Ravi B, Jenkinson R, Austin PC, *et al.* Relation between surgeon volume and risk of complications after total hip arthroplasty: propensity score matched cohort study. *BMJ* 2014; **348**: g3284.

538 Smabrekke A, Espehaug B, Havelin LI, Furnes O. Operating time and survival of primary total hip replacements: an analysis of 31,745 primary cemented and uncemented total hip replacements from local hospitals reported to the Norwegian Arthroplasty Register 1987-2001. *Acta Orthop Scand* 2004; **75**: 524–32.

539 Austin PC, Allignol A, Fine JP. The number of primary events per variable affects estimation of the subdistribution hazard competing risks model. *J Clin Epidemiol* 2017. DOI:10.1016/j.jclinepi.2016.11.017.

540 Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. *J Clin Epidemiol* 2015. DOI:10.1016/j.jclinepi.2014.12.014.

541 Zhang X, Zhang MJ, Fine J. A proportional hazards regression model for the subdistribution with right-censored and left-truncated competing risks data. *Stat Med* 2011; **30**: 1933–51.

542 Statacorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp

LLC., 2017.

543 Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: Sensitivity and specificity. *BMJ* 1994. DOI:10.1136/bmj.308.6943.1552.

544 Altman DG, Bland j. M. Statistics Notes: Diagnostic tests 2: Predictive values. *BMJ* 1994. DOI:10.1136/bmj.309.6947.102.

545 Altman DG, Bland JM. Statistics Notes: Diagnostic tests 3: Receiver operating characteristic plots. *BMJ* 1994. DOI:10.1136/bmj.309.6948.188.

546 Li F, He H. Assessing the Accuracy of Diagnostic Tests. *Shanghai Arch psychiatry* 2018; **30**: 207–12.

547 Ranstam J, Kärrholm J, Pulkkinen P, *et al*. Statistical analysis of arthroplasty data. I. Introduction and background. *Acta Orthop* 2011. DOI:10.3109/17453674.2011.588862.

548 Bewick V, Cheek L, Ball J. Statistics review 12: Survival analysis. Crit. Care. 2004. DOI:10.1186/cc2955.

549 Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009; **170**: 244–56.

550 Andersen PK, Geskus RB, De witte T, Putter H. Competing risks in epidemiology: Possibilities and pitfalls. *Int J Epidemiol* 2012. DOI:10.1093/ije/dyr213.

551 Stata. stcrreg — Competing-risks regression. *Stata Manuals* https://www.stata.com/manuals13/ststcrreg.pdf.

552 Penfold CM, Blom AW, Sayers A, *et al.* Understanding the uptake of new hip replacement implants in the UK: A cohort study using data from the National Joint Registry for England and Wales. *BMJ Open* 2019. DOI:10.1136/bmjopen-2019-029572.

553 Jimenez-Garca R, Villanueva-Martnez M, Fernandez-De-Las-Penas C, *et al.* Trends in primary total hip arthroplasty in Spain from 2001 to 2008: Evaluating changes in demographics, comorbidity, incidence rates, length of stay, costs and mortality. *BMC Musculoskelet Disord* 2011. DOI:10.1186/1471-2474-12-43.

554 Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg. Med. J. 2003. DOI:10.1136/emj.20.5.453.

555 Royston P, Barthel FMS. Projection of power and events in clinical trials with a time-to-event outcome. *Stata J* 2010. DOI:10.1177/1536867x1001000306.

556 National Joint Registry. 15th Annual Report 2018. *15th Annu Rep* 2018. DOI:10.1038/nmat2505.

557 Seppanen M, Laaksonen I, Pulkkinen P, *et al.* High revision rate for large-head metal-on-metal THA at a mean of 7.1 Years:A registry study. *Clin Orthop Relat Res* 2018. DOI:10.1007/s11999.0000000000000159.

558 Huang DCT, Tatman P, Mehle S, Gioe TJ. Cumulative revision rate is higher in metal-on-metal THA than metal-on-polyethylene THA: Analysis of survival in a community registry. *Clin Orthop Relat Res* 2013. DOI:10.1007/s11999-013-2821-1.

559 Bottle A, Parikh S, Aylin P, Loeffler M. Risk factors for early revision after total hip and knee arthroplasty: National observational study from a surgeon and

population perspective. *PLoS One* 2019. DOI:10.1371/journal.pone.0214855.

560 Ras Sørensen SAL, Jørgensen HL, Sporing SL, Lauritzen JB. Revision rates for metal-on-metal hip resurfacing and metal-on-metal total hip arthroplasty - A systematic review. HIP Int. 2016. DOI:10.5301/hipint.5000444.

561 Domínguez-Berjón M, Borrell C, Serral G, *et al.* Constructing a deprivation index based on census data in large Spanish cities (the MEDEA Project). *Gac Sanit* 2008; **22**: 179–87.

562 Reyes C, Garcia-Gil M, Elorza JM, *et al.* Socio-economic status and the risk of developing hand, hip or knee osteoarthritis: A region-wide ecological study. *Osteoarthr Cartil* 2015. DOI:10.1016/j.joca.2015.03.020.

563 Siddique J, Harel O, Crespi CM, Hedeker D. Binary variable multiple-model multiple imputation to address missing data mechanism uncertainty: Application to a smoking cessation trial. *Stat Med* 2014. DOI:10.1002/sim.6137.

564 NHS. Cardiovascular disease. https://www.nhs.uk/conditions/cardiovascular-disease/ (accessed May 5, 2020).

565 World Health Organization. Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed May 5, 2020).

566 NHS Digital. Finalised Patient Reported Outcome Measures (PROMs) in England for Hip & Knee Replacements, April 2018 – March 2019. 2020 https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/finalised-hip--knee-replacements-april-

2018---march-2019.

567 Strom BL. Overview of Automated Databases in Pharmacoepidemiology. In: Textbook of Pharmacoepidemiology. 2013. DOI:10.1002/9781118344828.ch08.

568 McAlindon TE, Bannuru RR, Sullivan MC, *et al.* OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthr Cartil* 2014. DOI:10.1016/j.joca.2014.01.003.

569 Wettermark B, Zoëga H, Furu K, *et al.* The nordic prescription databases as a resource for pharmacoepidemiological research-a literature review. Pharmacoepidemiol. Drug Saf. 2013. DOI:10.1002/pds.3457.

570 Lee TA, Bartle B, Weiss KB. Impact of NSAIDS on Mortality and the Effect of Preexisting Coronary Artery Disease in US Veterans. *Am J Med* 2007. DOI:10.1016/j.amjmed.2006.03.003.

571 Cooper C, Chapurlat R, Al-Daghri N, *et al.* Safety of Oral Non-Selective Non-Steroidal Anti-Inflammatory Drugs in Osteoarthritis: What Does the Literature Say? Drugs and Aging. 2019. DOI:10.1007/s40266-019-00660-1.

572 Chou R, McDonagh M, Nakamoto E, Griffin J. Analgesics for Osteoarthritis: An Update of the 2006 Comparative Effectiveness Review. *Agency fo Healthc Res Qual* 2011.

573 Hochberg MC. Mortality in osteoarthritis. *Clin Exp Rheumatol* 2008.

574 Nüesch E, Dieppe P, Reichenbach S, Williams S, Iff S, Jüni P. All cause and disease specific mortality in patients with knee or hip osteoarthritis: Population based cohort study. *BMJ* 2011. DOI:10.1136/bmj.d1165.

575 Cooper C, Arden NK. Excess mortality in osteoarthritis. BMJ. 2011. DOI:10.1136/bmj.d1407.

576 Hawker GA, Croxford R, Bierman AS, *et al.* All-cause mortality and serious cardiovascular events in people with hip and knee osteoarthritis: A population based cohort study. *PLoS One* 2014. DOI:10.1371/journal.pone.0091286.

577 Kluzek S, Sanchez-Santos MT, Leyland KM, *et al.* Painful knee but not hand osteoarthritis is an independent predictor of mortality over 23 years follow-up of a population-based cohort of middle-aged women. *Ann Rheum Dis* 2016. DOI:10.1136/annrheumdis-2015-208056.

578 Inacio MCS, Cashman K, Pratt NL, *et al.* Prevalence and changes in analgesic medication utilisation 1 year prior to total joint replacement in an older cohort of patients. *Osteoarthr Cartil* 2018. DOI:10.1016/j.joca.2017.11.016.

579 Jørgensen CC, Petersen M, Kehlet H, Aasvang EK. Analgesic consumption trajectories in 8975 patients 1 year after fast-track total hip or knee arthroplasty. *Eur J Pain (United Kingdom)* 2018. DOI:10.1002/ejp.1232.

580 Abrahams MJ. Neuropathic pain in soft tissue complaints. Best Pract. Res. Clin. Rheumatol. 2007. DOI:10.1016/j.berh.2006.12.005.

581 Hochman JR, Davis AM, Elkayam J, Gagliese L, Hawker GA. Neuropathic pain symptoms on the modified painDETECT correlate with signs of central sensitization in knee osteoarthritis. *Osteoarthr Cartil* 2013. DOI:10.1016/j.joca.2013.06.023.

582 Hochman JR, Gagliese L, Davis AM, Hawker GA. Neuropathic pain symptoms

in a community knee OA cohort. *Osteoarthr Cartil* 2011. DOI:10.1016/j.joca.2011.03.007.

583 Murphy SL, Lyden AK, Phillips K, Clauw DJ, Williams DA. Subgroups of older adults with osteoarthritis based upon differing comorbid symptom presentations and potential underlying pain mechanisms. *Arthritis Res Ther* 2011. DOI:10.1186/ar3449.

584 Maeda K, Sonohata M, Kitajima M, Kawano S, Mawatari M. Risk Factors of Neuropathic Pain after Total Hip Arthroplasty. *Hip Pelvis* 2018. DOI:10.5371/hp.2018.30.4.226.

585 Birke H, Ekholm O, Sjøgren P, Kurita GP, Højsted J. Long-term opioid therapy in Denmark: A disappointing journey. *Eur J Pain (United Kingdom)* 2017. DOI:10.1002/ejp.1053.

586 Braden JB, Fan MY, Edlund MJ, Martin BC, DeVries A, Sullivan MD. Trends in Use of Opioids by Noncancer Pain Type 2000-2005 Among Arkansas Medicaid and HealthCore Enrollees: Results From the TROUP Study. *J Pain* 2008. DOI:10.1016/j.jpain.2008.06.002.

587 von Korff M, Kolodny A, Deyo RA, Chou R. Long-term opioid therapy reconsidered. *Ann Intern Med* 2011. DOI:10.7326/0003-4819-155-5-201109060-00011.

588 Inacio MCS, Hansen C, Pratt NL, Graves SE, Roughead EE. Risk factors for persistent and new chronic opioid use in patients undergoing total hip arthroplasty: A retrospective cohort study. *BMJ Open* 2016. DOI:10.1136/bmjopen-2015-010664.

589 Namba RS, Inacio MCS, Pratt NL, Graves SE, Roughead EE, Paxton EW. Persistent Opioid Use Following Total Knee Arthroplasty: A Signal for Close Surveillance. *J Arthroplasty* 2018. DOI:10.1016/j.arth.2017.09.001.

590 Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015. DOI:10.1093/ije/dyv098.

591 Leahy TP, Ramagopalan S, Sammon C. The use of UK primary care databases in health technology assessments carried out by the National Institute for health and care excellence (NICE). *BMC Health Serv Res* 2020; **20**: 675.

592 Trimba R, Laughlin RT, Krishnamurthy A, Ross JS, Fox JP. Hospital-Based Acute Care After Total Hip and Knee Arthroplasty: Implications for Quality Measurement. *J Arthroplasty* 2016. DOI:10.1016/j.arth.2015.10.019.

593 Kelly MP, Prentice HA, Wang W, Fasig BH, Sheth DS, Paxton EW. Reasons for Ninety-Day Emergency Visits and Readmissions After Elective Total Joint Arthroplasty: Results From a US Integrated Healthcare System. *J Arthroplasty* 2018. DOI:10.1016/j.arth.2018.02.010.

594 Saleh A, Faour M, Sultan AA, Brigati DP, Molloy RM, Mont MA. Emergency Department Visits Within Thirty Days of Discharge After Primary Total Hip Arthroplasty: A Hidden Quality Measure. *J Arthroplasty* 2019. DOI:10.1016/j.arth.2018.08.032.

595 Zhang W, Nuki G, Moskowitz RW, *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. *Osteoarthr Cartil* 2010. DOI:10.1016/j.joca.2010.01.013.

596 Wartemberg GK, Goff T, Jones S, Newman J. Extended Routine Follow-up of Primary Uncomplicated Hip and Knee Arthroplasty is not Necessary. *Open Orthop J* 2020; **14**. DOI:10.2174/1874325001913010266.

597 Partridge CJ. The Canterbury study. Physiotherapy in the community. *Physiotherapy* 1977.

598 Hawker GA, Guan J, Croxford R, *et al.* A prospective population-based study of the predictors of undergoing total joint arthroplasty. *Arthritis Rheum* 2006. DOI:10.1002/art.22146.

599 Czoski Murray CJ, Kingsbury SR, Arden NK, *et al.* Towards UK poSt Arthroplasty Follow-up rEcommendations (UK SAFE): Protocol for an evaluation of the requirements for arthroplasty follow-up, and the production of consensus-based recommendations. *BMJ Open* 2019. DOI:10.1136/bmjopen-2019-031351.

600 Briggs T. Getting it right first time. 2012 offering patients more expensive implants with little or no added benefit denies other patients Orthopaedic care.

601 Smith LK. A Survey of the Current State of Hip Arthroplasty Surveillance in the United Kingdom. *Musculoskeletal Care* 2014. DOI:10.1002/msc.1077.

602 Brekke KR, Nuscheler R, Straume OR. Gatekeeping in health care. *J Health Econ* 2007. DOI:10.1016/j.jhealeco.2006.04.004.

603 Li Y, Salmasian H, Harpaz R, Chase H, Friedman C. Determining the reasons for medication prescriptions in the EHR using knowledge and natural language

processing. *AMIA Annu Symp Proc* 2011.

604 Menendez ME, Ring D, Bateman BT. Preoperative Opioid Misuse is Associated With Increased Morbidity and Mortality After Elective Orthopaedic Surgery. *Clin Orthop Relat Res* 2015; **473**: 2402–12.

605 Koffeman AR, Valkhoff VE, Çelik S, *et al.* High-risk use of over-the-counter non-steroidal anti-inflammatory drugs: A population-based cross-sectional study. *Br J Gen Pract* 2014. DOI:10.3399/bjgp14X677815.

606 Trancik T, Mills W, Vinson N. The effect of indomethacin, aspirin, and ibuprofen on bone ingrowth into a porous-coated implant. In: Clinical Orthopaedics and Related Research. 1989.

607 Ribeiro F V., César-Neto JB, Nociti FH, *et al.* Selective Cyclooxygenase-2 Inhibitor May Impair Bone Healing Around Titanium Implants in Rats. *J Periodontol* 2006. DOI:10.1902/jop.2006.060119.

608 Allami MK, Giannoudis P V. Cox inhibitors and bone healing. Acta Orthop. Scand. 2003. DOI:10.1080/00016470310018351.

609 Jacobsson SA, Djerf K, Ivarsson I, Wahlstrom O. Effect of diclofenac on fixation of hydroxyapatite-coated implants: An experimental study. *J Bone Jt Surg - Ser B* 1994.

610 Gerstenfeld LC, Einhorn TA. COX inhibitors and their effects on bone healing. Expert Opin. Drug Saf. 2004. DOI:10.1517/14740338.3.2.131.

611 Nizam I, Kohan L, Field C, Kerr D. Do nonsteroidal anti-inflammatory drugs cause endoprosthetic loosening? Mid- to long-term follow-up of 100 total hip

arthroplasties after local NSAID infiltration. *Biomed Res Int* 2015. DOI:10.1155/2015/703071.

612 Zhang AH, Chen X, Zhao QX, Wang KL. A systematic review and meta-analysis of naproxen for prevention heterotopic ossification after hip surgery. *Medicine (Baltimore)* 2019. DOI:10.1097/MD.0000000000014607.

613 P. Kjaersgaard-Andersen, S. Johnsen, S. Overgaard, A. Petersen AR. NSAIDs reduced the risk of revision due aseptic loosening in cemented Total Hip Replacement - Results from the Danish Hip Arthroplasty Register. *Orthop Proc* 2010; **92-B**: 522.

614 UK Government. Pregabalin and gabapentin to be controlled as class C drugs. 2018. https://www.gov.uk/government/news/pregabalin-and-gabapentin-to-be-controlled-as-class-c-drugs (accessed June 17, 2020).

615 Kushnir T, Greenberg D, Madjar N, Hadari I, Yermiahu Y, Bachner YG. Is burnout associated with referral rates among primary care physicians in community clinics. *Fam Pract* 2014. DOI:10.1093/fampra/cmt060.

616 Hobbs FDR, Bankhead C, Mukhtar T, *et al.* Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *Lancet* 2016. DOI:10.1016/S0140-6736(16)00620-6.

617 Glazier RH, Dalby DM, Badley EM, Hawker GA, Bell MJ, Buchbinder R. Determinants of physician confidence in the primary care management of musculoskeletal disorders. *J Rheumatol* 1996.

618 Matheny JM, Brinker MR, Elliott MN, Blake R, Rowane MP. Confidence of

graduating family practice residents in their management of musculoskeletal conditions. *Am J Orthop (Belle Mead NJ)* 2000.

619 Roland MO, Porter RW, Matthews JG, Redden JF, Simonds GW, Bewley B. Improving care: A study of orthopaedic outpatient referrals. *Br Med J* 1991. DOI:10.1136/bmj.302.6785.1124.

620 O'Cathain A, Froggett M, Taylor MP. General practice based physiotherapy: Its use and effect on referrals to hospital orthopaedics and rheumatology outpatient departments. *Br J Gen Pract* 1995.

621 Carter RH, Densley JA, Galley CM, Holland A, Jones LE, Dunn CDR. Factors associated with GP referrals to physiotherapy. *Int J Ther Rehabil* 2013. DOI:10.12968/bjtr.2001.8.12.13820.

622 Clemence ML, Seamark DA. GP referral for physiotherapy to musculoskeletal conditions—a qualitative study. *Fam Pract* 2003; **20**: 578–82.

623 Wu AW, Snyder C, Clancy CM, Steinwachs DM. Adding the patient perspective to comparative effectiveness research. *Health Aff* 2010. DOI:10.1377/hlthaff.2010.0660.

624 The National Joint Registry of England and Wales. The National Joint Registry 18th Annual Report 2021. 2021 https://reports.njrcentre.org.uk/Portals/0/PDFdownloads/NJR 18th Annual Report 2021.pdf.

625 Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999; **319**.

DOI:10.1136/bmj.319.7223.1492.

626 Petrie A. Statistics in orthopaedic papers. *J Bone Joint Surg Br* 2006; **88-B**. DOI:10.1302/0301-620X.88B9.17896.

627 Stata Manual. Power and sample size for the Cochran–Mantel–Haenszel test. https://www.stata.com/manuals/pss-2powercmh.pdf.

628 Graves SE, Rothwell A, Tucker K, Jacobs JJ, Sedrakyan A. A multinational assessment of metal-on-metal bearings in hip replacement. *J Bone Joint Surg Am* 2011. DOI:10.2106/JBJS.K.01220.

629 Podmore B, Hutchings A, Skinner JA, MacGregor AJ, van der Meulen J. Impact of comorbidities on the safety and effectiveness of hip and knee arthroplasty surgery. *Bone Joint J* 2021; **103-B**: 56–64.

630 Malak TT, Broomfield JAJ, Palmer AJR, *et al*. Surrogate markers of long-term outcome in primary total hip arthroplasty: A systematic review. Bone Jt. Res. 2016. DOI:10.1302/2046-3758.56.2000568.

631 Ranstam J, Kärrholm J, Pulkkinen P, *et al*. Statistical analysis of arthroplasty data. II. Guidelines. *Acta Orthop* 2011. DOI:10.3109/17453674.2011.588863.

632 Van Der Pas S, Nelissen R, Fiocco M. Different competing risks models for different questions may give similar results in arthroplasty registers in the presence of few events: Illustrated with 138,234 hip (124,560 patients) and 139,070 knee (125,213 patients) replacements from the Dutch Arth. *Acta Orthop* 2018. DOI:10.1080/17453674.2018.1427314.

633 NHS. NSAIDs. 2019. https://www.nhs.uk/conditions/nsaids/ (accessed Aug

12, 2020).

634 Finley CR, Chan DS, Garrison S, *et al.* What are the most common conditions in primary care? Systematic review. *Can Fam Physician* 2018.

635 Pedersen AB, Mikkelsen EM, Cronin-Fenton D, *et al.* Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol* 2017. DOI:10.2147/CLEP.S129785.

636 Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression: Third Edition. 2013 DOI:10.1002/9781118548387.

637 Anand R, Graves SE, de Steiger RN, *et al.* What is the benefit of introducing new hip and knee prostheses? *J Bone Joint Surg Am* 2011. DOI:10.2106/JBJS.K.00867.

638 European Commission. Exploring Innovative Healthcare: The role of Medical Technology Innovation and Regulation. 2011.

639 Sorrel S. Applied Clinical trials online. Medical device development: US and EU differences. 2006.

640 Cafri G, Graves SE, Sedrakyan A, *et al.* Postmarket surveillance of arthroplasty device components using machine learning methods. *Pharmacoepidemiol Drug Saf* 2019. DOI:10.1002/pds.4882.

641 The National Joint Registry of England and Wales. Highlights: Component database re-design. 2019. https://reports.njrcentre.org.uk/Tracking-implant-performance (accessed Aug 15, 2020).

642 Hawker GA, Wright JG, Badley EM, Coyte PC. Perceptions of, and willingness

to consider, total joint arthroplasty in a population-based cohort of individuals with disabling hip and knee arthritis. *Arthritis Care Res (Hoboken)* 2004. DOI:10.1002/art.20524.

643 Bulstrode C. Keeping up with orthopaedic epidemics. *Br Med J (Clin Res Ed)* 1987. DOI:10.1136/bmj.295.6597.514.

644 Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* 1958. DOI:10.1080/01621459.1958.10501452.

645 National Joint Registry for England Wales Northern Ireland and the Isle of Man. 16th Annual Report 2019:National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. *NJR 16th Annu Rep 2019* 2019.

646 Lübbeke A, Rothman KJ, Garavaglia G, *et al.* Strong association between smoking and the risk of revision in a cohort study of patients with metal-on-metal total hip arthroplasty. *J Orthop Res* 2014. DOI:10.1002/jor.22603.

647 Gundtoft PH, Varnum C, Pedersen AB, Overgaard S. The Danish hip arthroplasty register. Clin. Epidemiol. 2016. DOI:10.2147/CLEP.S99498.

648 Kärrholm J. The Swedish Hip Arthroplasty Register (www.shpr.se). Acta Orthop. 2010. DOI:10.3109/17453671003635918.

649 Labek G, Schöffl H, Meglic M. New medical device regulations ahead - What does that mean for arthroplasty registers? Acta Orthop. 2015. DOI:10.3109/17453674.2014.1002185.

650 Black N, Tan S. Use of national clinical databases for informing and for evaluating health care policies. *Health Policy (New York)* 2013.

DOI:10.1016/j.healthpol.2012.10.007.

651 Brown JS, Kulldorff M, Chan KA, *et al.* Early detection of adverse drug events within population-based health networks: Application of sequential testing methods. *Pharmacoepidemiol Drug Saf* 2007. DOI:10.1002/pds.1509.

652 Oliveira JL, Lopes P, Nunes T, *et al.* The EU-ADR Web Platform: Delivering advanced pharmacovigilance tools. *Pharmacoepidemiol Drug Saf* 2013. DOI:10.1002/pds.3375.

653 Zywiel MG, Stroh DA, Lee SY, Bonutti PM, Mont MA. Chronic opioid use prior to total knee arthroplasty. *J Bone Jt Surg - Ser A* 2011. DOI:10.2106/JBJS.J.01473.

654 Lieu TA, Kulldorff M, Davis RL, *et al.* Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007. DOI:10.1097/mlr.0b013e3180616c0a.

655 Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value Heal* 2012. DOI:10.1016/j.jval.2012.06.015.

656 Rolka H, Walker DW, English R, *et al.* Analytical challenges for emerging public health surveillance. *MMWR Surveill Summ* 2012.

657 Franklin PD, Allison JJ, Ayers DC. Beyond joint implant registries: A patient-centered research consortium for comparative effectiveness in total joint replacement. JAMA - J. Am. Med. Assoc. 2012. DOI:10.1001/jama.2012.12568.

658 Sueyoshi T, Meding JB, Davis KE, Lackey WG, Malinzak RA, Ritter MA. Clinical predictors for possible failure after total hip arthroplasty. *HIP Int* 2016. DOI:10.5301/hipint.5000389.

659 Callahan A, Fries JA, Ré C, *et al.* Medical device surveillance with electronic health records. *npj Digit Med* 2019. DOI:10.1038/s41746-019-0168-z.

660 Lalmohamed A, MacGregor AJ, de Vries F, Leufkens HGM, van Staa TP. Patterns of Risk of Cancer in Patients with Metal-on-Metal Hip Replacements versus Other Bearing Surface Types: A Record Linkage Study between a Prospective Joint Registry and General Practice Electronic Health Records in England. *PLoS One* 2013. DOI:10.1371/journal.pone.0065891.

661 Suvarna VR. Real world evidence (RWE) - Are we (RWE) ready? Perspect. Clin. Res. 2018. DOI:10.4103/picr.PICR-36-18.

662 Fanaroff AC, Steffel J, Alexander JH, Lip GYH, Califf RM, Lopes RD. Stroke prevention in atrial fibrillation: Re-defining 'real-world data' within the broader data universe. Eur. Heart J. 2018. DOI:10.1093/eurheartj/ehy236.

663 Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. N. Engl. J. Med. 2010. DOI:10.1056/NEJMp1006114.

664 Harpaz R, Vilar S, DuMouchel W, *et al.* Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Informatics Assoc* 2013. DOI:10.1136/amiajnl-2012-000930.

665 EFORT. NORE Research & Publications. 2020. https://www.efort.org/about-us/nore/research/ (accessed Aug 22, 2020).

666 The National Joint Registry of England and Wales. 8th Annual Report. 2011.

667 Office of National Statistics. Overview of the UK population: November 2018. 2018.

https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/november2018 (accessed Aug 25, 2020).

668 Whitehouse SL, Bolland BJ, Howell JR, Crawford RW, Timperley AJ. Mortality following hip arthroplasty--inappropriate use of National Joint Registry (NJR) data. *J Arthroplast* 2014; **29**: 1827–34.

669 Bernal-Delgado E, Garcia-Armesto S, Oliva J, *et al.* Spain: Health System Review. Health Syst. Transit. 2018.

670 Raftery J. Review of NICE's recommendations, 1999-2005. *BMJ* 2006. DOI:10.1136/bmj.332.7552.1266.

671 Oleske DM, Bonafede MM, Jick S, Ji M, Hall JA. Electronic health databases for epidemiological research on joint replacements: Considerations when making cross-national comparisons. *Ann Epidemiol* 2014. DOI:10.1016/j.annepidem.2014.06.003.

672 Martinsen DS, Vasev N. Cross-border healthcare in the European Union: In: Studying public policy. 2017. DOI:10.2307/j.ctt1t898hb.29.

673 Singh JA, Kundukulam JA, Bhandari M. A systematic review of validated methods for identifying orthopedic implant removal and revision using administrative data. *Pharmacoepidemiol Drug Saf* 2012. DOI:10.1002/pds.2309.

674 Mnatzaganian G, Ryan, Norman P, Davidson, Hiller J. Length of stay in hospital and all-cause readmission following elective total joint replacement in elderly men. *Orthop Res Rev* 2012. DOI:10.2147/orr.s31668.

675 Sterne JAC, White IR, Carlin JB, *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj* 2009; **338**: 1–10.

676 Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006. DOI:10.1016/j.jclinepi.2006.01.014.

677 Little RJA. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc* 1988. DOI:10.1080/01621459.1988.10478722.

678 Jordan L. The problem with Big Data in Translational Medicine. A review of where we've been and the possibilities ahead. Appl. Transl. Genomics. 2015. DOI:10.1016/j.atg.2015.07.005.

679 Romero L, Nieuwenhuijse M, Carr A, Sedrakyan A. Review of clinical outcomes-based anchors of minimum clinically important differences in hip and knee registry-based reports and publications. J. Bone Jt. Surg. - Am. Vol. 2014. DOI:10.2106/JBJS.N.00464.

680 Ibrahim AM, Dimick JB. Monitoring medical devices: Missed warning signs within existing data. JAMA - J. Am. Med. Assoc. 2017. DOI:10.1001/jama.2017.6584.

681 Resnic FS, Matheny ME. Medical devices in the real world. *N Engl J Med* 2018. DOI:10.1056/NEJMp1712001.

682 De Steiger RN, Miller LN, Davidson DC, Ryan P, Graves SE. Joint registry approach for identification of outlier prostheses. *Acta Orthop* 2013. DOI:10.3109/17453674.2013.831320.

683 Biondi-Zoccai G, Romagnoli E, Agostoni P, *et al.* Are propensity scores really superior to standard multivariable analysis? Contemp. Clin. Trials. 2011. DOI:10.1016/j.cct.2011.05.006.

684 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. In: Matched Sampling for Causal Effects. 2006. DOI:10.1017/CBO9780511810725.016.

685 Rubin DB. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Stat Med* 2007. DOI:10.1002/sim.2739.

686 Austin PC. The use of propensity score methods with survival or time-to-event outcomes: Reporting measures of effect similar to those used in randomized experiments. *Stat Med* 2014. DOI:10.1002/sim.5984.

687 Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003. DOI:10.1093/aje/kwg115.

688 Krucoff MW, Sedrakyan A, Normand SLT. Bridging unmet medical device ecosystem needs with strategically coordinated registries networks. JAMA - J. Am. Med. Assoc. 2015. DOI:10.1001/jama.2015.11036.

689 Sedrakyan A, Campbell B, Graves S, Cronenwett JL. Surgical registries for advancing quality and device surveillance. Lancet. 2016. DOI:10.1016/S0140-6736(16)31402-7.

690 Zagra L. Advances in hip arthroplasty surgery: What is justified? *EFORT Open*

Rev 2017. DOI:10.1302/2058-5241.2.170008.

691 Ninomiya JT, Dean JC, Incavo SJ. What's new in hip replacement. *J Bone Jt Surg - Am Vol* 2017. DOI:10.2106/JBJS.17.00704.

692 Nieuwenhuijse MJ, Nelissen RGHH, Schoones JW, Sedrakyan A. Appraisal of evidence base for introduction of new implants in hip and knee replacement: A Systematic review of five widely used device technologies. BMJ. 2014. DOI:10.1136/bmj.g5133.

693 Howard JJ. Balancing innovation and medical device regulation: The case of modern metal-on-metal hip replacements. Med. Devices Evid. Res. 2016. DOI:10.2147/MDER.S113067.

694 Wood A m., White I r., Thompson S g. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin. Trials. 2004. DOI:10.1191/1740774504cn032oa.

695 van Buuren S. Flexible Imputation of Missing Data. 2012 DOI:10.1201/b11826.

696 Pedersen AB, Baggesen LM, Ehrenstein V, Pedersen L, Lasgaard M, Mikkelsen EM. Perceived stress and risk of any osteoporotic fracture. *Osteoporos Int* 2016. DOI:10.1007/s00198-016-3490-1.

697 Royston P. Multiple imputation of missing values: Update. *Stata J* 2005. DOI:10.1177/1536867x0500500204.

698 Royston P, White IR. Multiple imputation by chained equations (MICE): Implementation in Stata. *J Stat Softw* 2011. DOI:10.18637/jss.v045.i04.

699 Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons, LTD, 1987.

700 Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2002 DOI:10.1002/9781119013563.

701 Rubin DB. Inference and missing data. *Biometrika* 1976; **63**: 581–92.

702 Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am. J. Epidemiol. 1995. DOI:10.1093/oxfordjournals.aje.a117592.

703 Carpenter JR, Kenward MG. A critique of common approaches to missing data (&) MAR methods for quantitative data. *In; Missing data randomised Control trials - a Pract Guid* 2008.

704 Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol* 1991. DOI:10.1093/oxfordjournals.aje.a116164.

705 Pedersen AB, Sorensen HT, Mehnert F, Overgaard S, Johnsen SP. Risk factors for venous thromboembolism in patients undergoing total hip replacement and receiving routine thromboprophylaxis. *J Bone Jt Surg - Ser A* 2010. DOI:10.2106/JBJS.I.00882.

706 Barnes SA, Larsen MD, Schroeder D, Hanson A, Decker PA. Missing data assumptions and methods in a smoking cessation study. *Addiction* 2010. DOI:10.1111/j.1360-0443.2009.02809.x.

707 Goldstein H, Carpenter J, Kenward MG, Levin KA. Multilevel models with multivariate mixed response types. *Stat Modelling* 2009. DOI:10.1177/1471082X0800900301.

708 Knorr-Held L. Analysis of Incomplete Multivariate Data. J. L. Schafer, Chapman & Hall, London, 1997. *Stat Med* 2000; **19**: 1006–8.

709 Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *J Am Stat Assoc* 1999. DOI:10.1080/01621459.1999.10473862.

710 Wood AM, White IR, Hillsdon M, Carpenter J. Comparison of imputation and modelling methods in the analysis of a physical activity trial with missing outcomes. Int. J. Epidemiol. 2005. DOI:10.1093/ije/dyh297.

711 Moodie EEM, Delaney JAC, Lefebvre G, Platt RW. Missing confounding data in marginal structural models: A comparison of inverse probability weighting and multiple imputation. *Int J Biostat* 2008. DOI:10.2202/1557-4679.1106.

712 Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. Am. J. Epidemiol. 2008. DOI:10.1093/aje/kwn071.

713 Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006. DOI:10.1016/j.jclinepi.2006.01.009.

714 Rubin DB. Multiple Imputation after 18+ Years. *J Am Stat Assoc* 1996. DOI:10.1080/01621459.1996.10476908.

715 Carpenter JR, Kenward MG. Multiple Imputation and its Application. 2012

DOI:10.1002/9781119942283.

716 Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: A simulation study. *BMC Med Res Methodol* 2010. DOI:10.1186/1471-2288-10-7.

717 Ali AMG, Dawson SJ, Blows FM, *et al.* Comparison of methods for handling missing data on immunohistochemical markers in survival analysis of breast cancer. *Br J Cancer* 2011. DOI:10.1038/sj.bjc.6606078.

718 Martín-Merino E, Calderón-Larrañaga A, Hawley S, *et al.* The impact of different strategies to handle missing data on both precision and bias in a drug safety study: A multidatabase multinational population-based cohort study. *Clin Epidemiol* 2018. DOI:10.2147/CLEP.S154914.

719 Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999. DOI:10.1177/096228029900800102.

720 Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol* 2019. DOI:10.1016/j.jclinepi.2019.02.016.

721 Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. *BMC Med Res Methodol* 2017. DOI:10.1186/s12874-017-0442-1.

722 McNeish D. Missing data methods for arbitrary missingness with small samples. *J Appl Stat* 2017. DOI:10.1080/02664763.2016.1158246.

723 Clavel J, Merceron G, Escarguel G. Missing data estimation in morphometrics: How much is too much? *Syst Biol* 2014. DOI:10.1093/sysbio/syt100.

724 Lee J, Huber JJ. Multiple imputation with large proportions of missing data: How much is too much? United Kingdom Stata Users' Gr. Meet. 2011. 2011.

725 Lee KJ, Carlin JB. Recovery of information from multiple imputation: A simulation study. Emerg. Themes Epidemiol. 2012. DOI:10.1186/1742-7622-9-3.

726 De Silva AP, Moreno-Betancur M, De Livera AM, Lee KJ, Simpson JA. Multiple imputation methods for handling missing values in a longitudinal categorical variable with restrictions on transitions over time. *BMC Med Res Methodol* 2019. DOI:10.1186/s12874-018-0653-0.

727 De Silva AP, Moreno-Betancur M, De Livera AM, Lee KJ, Simpson JA. A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: A simulation study. *BMC Med Res Methodol* 2017. DOI:10.1186/s12874-017-0372-y.

728 Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *Int J Methods Psychiatr Res* 2011. DOI:10.1002/mpr.329.

729 Sullivan TR, Lee KJ, Ryan P, Salter AB. Multiple imputation for handling missing outcome data when estimating the relative risk. *BMC Med Res Methodol* 2017. DOI:10.1186/s12874-017-0414-5.

730 Graham JW, Olchowski AE, Gilreath TD. How many imputations are really

needed? Some practical clarifications of multiple imputation theory. *Prev Sci* 2007. DOI:10.1007/s11121-007-0070-9.

731 White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011. DOI:10.1002/sim.4067.

732 Bodner TE. What improves with increased missing data imputations? *Struct Equ Model* 2008. DOI:10.1080/10705510802339072.

733 Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol* 2015; **15**: 30.

734 White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; **30**: 377–99.

735 Rothman K. Epidemiology: An Introduction. New York: Oxford University Press, 2002.

736 Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. Second Edition of the ERIC Notebook: Risk and Rate Measures in Cohort Studies. *ERIC Noteb* 2015.

737 Kleinbaum DGD, Klein M. Survival Analysis: A Self-Learning Text, Third Edition (Statistics for Biology and Health). 2011 DOI:10.1016/B978-0-12-387667-6.00013-0.

738 Cutler SJ, Ederer F. Maximum utilization of the life table method in analyzing survival. *J Chronic Dis* 1958. DOI:10.1016/0021-9681(58)90126-7.

739 Mantel N, Haenszel W. Statistical aspects of the analysis of data from

retrospective studies of disease. *J Natl Cancer Inst* 1959. DOI:10.1093/jnci/22.4.719.

740 Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc* 1999. DOI:10.1080/01621459.1999.10474144.

741 Biau DJ, Latouche A, Porcher R. Competing events influence estimated survival probability: When is Kaplan-Meier analysis appropriate? Clin. Orthop. Relat. Res. 2007. DOI:10.1097/BLO.0b013e3180986753.

742 Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Stat Med* 1999. DOI:10.1002/(SICI)1097-0258(19990330)18:6<695::AID-SIM60>3.0.CO;2-O.

743 Wainwright C, Theis J-CC, Garneti N, Melloh M. Age at hip or knee joint replacement surgery predicts likelihood of revision surgery. *J Bone Jt Surg Br* 2011; **93**: 1411–5.

7. APPENDICIES

7.1. Appendix 1

Data Extraction Tool



Tracking Information

Article title: _____

_____

Year: _____

Paper ID: _____

Reviewer name: _____

Data

Study design

☐ Systematic Review ☐ Case Series

☐ Randomised Control Trial ☐ Anecdotal/Non-clinical

☐ Observational Study (include type): _____

Does the study meet inclusion/exclusion criteria? ☐ Yes ☐ No

- ☐ Investigate a surrogate marker measured post-surgery
- ☐ Surrogate marker is compared to revision rate as outcome
- ☐ Revision rate is at least 10 years (outcome)
- ☐ Primary THA investigated
- ☐ Indication for surgery is Osteoarthritis
- ☐ Inferential statistics is used to investigate association
- ☐ Exclude: Animal, non-english studies

What is the surrogate marker (intervention)? _____

Length of Follow up: _____

Sample size: _____

7.2. Appendix 2

Introduction

In epidemiological studies using observational data, and particularly with analysis of registry or primary care databases, missing data are unavoidable[675] and pervasive within clinical epidemiology[635]. The issue of missing data is not isolated to registry databases but various data sources and in multiple study designs[635]. There can be a number of reasons for missing data to occur including lack of measurement taking place, a patient may fail to attend a clinical appointment, and loss to follow-up among others[635]. Unfortunately, if missing data is not dealt with, the validity of the findings from a study can be undermined[694,695]. Patients with missing data may have systematic differences compared to patients with complete data in relation to the outcome of interest[635]. An example of this could be that healthier patients are less likely to attend a primary care appointment for their weight measurement and would thus be less likely to have their weight recorded. A common variable with missing data is smoking status and patients with missing values have been shown to have missing data in other lifestyle variables[696].

The most common method to deal with missing values is to perform a complete case analysis[675]. In this case, patients with missing values are excluded and only patients with no missing data are analysed. In addition to causing bias, in complete case analysis there can be a cumulative effect of excluding significant proportions of the original sample when there are several variables with missing data[675]. Therefore, the study will suffer from a loss of power and precision following analysis. While concerns of missing data has previously largely been ignored[694], advances in recent decades in

technology and computing power have enabled researchers to deal with missing data through multiple imputation more readily[675,697,698]. Multiple imputation methods offer increased precision and can reduce bias[675].

Types of missing data

How missing data may influence results and introduce bias depends on the type of missingness that exists. Types of missing data has been categorised by Rubin[699–701] according to their relation to the observed values and other variables, and thus giving insight to how they may be managed (Table 7-1). Firstly, missing data can be Missing Completely at Random (MCAR). In this scenario the missing values have no systematic differences between missing values and the observed values of the variable in question or other variables. An example would be the lack of a reading for the patient's weight due to the machine being broken at the time the recording was taking place and subsequently not measured and recorded. The second category is Missing at Random (MAR). Here, a systematic difference between the missing values and observed values exist which can be explained by the observed data. For example, the true weight may be higher for missing values compared to other measured values, due to factors such as being middle aged or male gender which are risk factors for obesity.

Finally, the third category is Missing not at Random (MNAR). When missing values are MNAR, a systematic difference remains when observed data are taken into account and are thus dependant on the missing values. Therefore, it is not possible to confidently make conclusions about the missing value based on the observed data. For

example, patients with a lower or higher weight are more likely to have their weight recorded compared to patients with normal weight even after adjusting for age.

Table 7-1: Rubin types of missing data

Type	Definition
Missing Completely at Random (MCAR)	No systematic difference between missing values and observed values
Missing at Random (MCR)	A systematic difference between missing values and observed values which can be explained by the observed data
Missing not at Random (MNAR)	A systematic difference between missing values and observed values which cannot be fully explained by the observed data

Rubin's categories are widely used. Complete case analysis remains the most commonly used method for addressing missing data[635] and is likely popular due to its ease of use.

Complete case analysis will likely lead to bias if missing data within a cohort can be plausibly categorised as 'missing at random' and not 'missing completely at random'[675]. To address this bias, techniques such as multiple imputation have been devised to allow the researcher to include patients with missing data within the overall analysis[675].

Methods of dealing with missing data

In epidemiological studies, the vast majority of missing data can be categorised as MAR and are not MCAR or MNAR[635]. Using this principle, there is a wide variety of techniques to address missing data allowing for imputation of new values to ensure the patients can be included for analysis.

The 'missing indicator method' is not a full imputation method but is used in epidemiological research. Here categorical variables create a missing group and for continuous variables missing values are assigned a fixed value such as zero[635]. However, even if there were minimal missing data this method is prone to bias[702]. In particular, this study has missing values in confounding variables, using missing indicator method would lead to residual confounding[635].

The next group of methods is single value imputation. One of the simplest, known as the 'last value carried forward', is to replace the missing value with the last value measured within the variable[703]. Another simple method is to replace missing values with the mean or median of the observed values[704]. A more advanced method would be to use simple regression-based single imputation where imputed values are predicted from observed values[635]. Unfortunately, these approaches can lead to significant bias and may render results invalid[675]. Research investigating such approaches have largely determined them to be statistically invalid[703,704], as standard errors are invariably too small when single imputation of missing values is performed.

This is due to such approaches failing to address the uncertainty of the missing values true measure. Therefore, results are in danger of having a Type 1 error[702].

Other methods are used in trials where either missing data is in a binary outcome or limited to the exposure variable.

Studies have employed a 'best-case' or 'worst-case' scenario approach to missing data by replacing missing values with the clinically best or worst value from observed data[705]. An example would be to replace missing values with the highest or lowest observed value and performing analysis on the two datasets separately. The results can then be compared. Unfortunately, interpretation can be challenging particularly if opposing results were yielded. This method is more commonly used in studies where missing data was in the outcome variable and more clinically meaningful 'best-case' or 'worst-case' scenario judgement can be made. However, studies have shown this method can provide biased estimates[706].

Compared to these methods, a statistically more powerful analysis can be performed to minimise bias if an assumption can be made of the missing data falling into the category of missing at random. This is achieved by methods taking into account the uncertainty of a missing value and using the observed data to impute missing values, thus allowing analysis to include patients with missing data. Again there are several methods to achieve this goal including random effects modelling using data from the partially observed variable at a specific time-point[707]. Bayesian methods have been employed by using data from the partially observed variables[708]. Finally, other

approaches such as maximum likelihood estimation[708] and weighted analysis have been used[709]. However, Multiple imputation has grown in popularity as it has shown to be more valid with less bias compared to these other approaches[635,710,711].

Multiple imputation

Multiple Imputation (MI) is now regarded as the standard method to address missing data in a variety of research areas[635]. Perhaps one of the main reasons for its rise in popularity, despite its computational complexity, is the availability of statistical software where MI is now relatively simple to perform[675]. The aim of MI is to account for the uncertainty surrounding missing values by creating multiple separate datasets with plausible imputed values using observed data and combining the results analysed in each[675]. In this manner, the issue of standard errors being either "too small or too large" can be solved[712]. In addition, MI aims to address the potential bias from other methods when investigating associations derived from observed data (i.e. obtaining results similar to those that would be calculated on a full dataset)[695].

One of the main assumptions of multiple imputation is that observed data can inform missing data and thus missing data are categorised as MAR[635,675]. While it is not possible to definitively ascertain whether missing data is MAR, multiple imputation can address missingness that is MCAR and MNAR[695]. Furthermore, the MAR assumption is not only plausible in most epidemiological research[635], increasing plausibility is obtained with the more variables included within the multiple imputation model[713]. A full description of the multiple imputation methods and

implementation is beyond the scope of this thesis. However, there are three key stages and certain factors to take into account that are addressed here.

The first step is to select the variables to incorporate within the multiple imputation model that could plausibly impute missing data. As a general principle, all the variables that will be used in the subsequent analysis regression should be included[635]. Any other variable that could inform the imputation model, known as auxiliary variables, should also be included[635]. Once these variables are identified, multiple imputed datasets are created. Missing values are imputed using a Bayesian approach with the predictive distribution of the observed data[675,714]. A variation of imputed values exists within each database to account for the uncertainty of the missing data as it is not possible to definitely know the true measure[635,675].

The second step is to run the standard statistical method to model association on each imputed dataset. Again, due to the variability following imputation within each dataset accounting for uncertainty, the estimations will also differ between each dataset. This will be important information subsequently used[713]. Each dataset calculation will have its own coefficients and standard errors as a measure of the associations.

The final step is to combine the estimates of association from each dataset using Rubin's rules[715]. The coefficients and corresponding standard errors resulting in

confidence intervals are also combined, and therefore the final estimate accounts for variation within datasets.

The minimum proportion of missing data in the variables for this study is 25.1%. Where a proportion of approximately 25% or more is missing the loss of precision using complete case analysis warrants utilising an alternative approach[716]. Thus, multiple imputation was used in the subsequent experimental chapters.

Implementation of multiple imputation

There are number of considerations to account for when constructing the multiple imputation model.

Which variables to include

The first consideration is which variables to include within the imputation model. As mentioned previously, all variables used for the subsequent analysis of association should be included[635]. In addition, 'auxiliary variables' should be investigated and incorporated if appropriate to reduce bias and increase precision[635]. For a variable to be considered as an auxiliary it should be associated with missing values or be associated with both the missing value and the probability of the data being missing[715]. If an auxiliary variable has association with both the missing values and the likelihood of missingness, they are more likely to impact subsequent imputation reducing bias[715]. In

the majority of studies, a priori variables both in the analytical model and potential auxiliary, are compiled using the research question, knowledge, and literature[635].

One of the potential pitfalls of multiple imputation is to construct a model that does not include the outcome of interest[675]. Studies have shown multiple imputation to perform better when the outcome of interest is included[717]. For time to event outcomes, it is important to include both whether a patient had the outcome of interest and follow up time using the Nelson-Aalen method.

In the studies of the subsequent experimental chapters of this thesis, all variables (exposures, confounders, and outcome) were included within the imputation model here. In addition, the Nelson-Aalen estimate was included[718]. Furthermore, all variables within the cohort were investigated to identify predictors of missingness by fitting exploratory regression models for each individual variable with missing data (separate models used for smoking missingness, one for alcohol missingness and one for BMI missingness). A backward stepwise selection method (with a p-exit<0.1 and p-entry<0.05) using an inbuilt Stata command was performed to identify only the key variables[718]. Additional variables included pharmaceutical use and other co-morbidities.

Proportion of missingness

There have been studies investigating what proportion of missingness would need multiple imputation to ensure unbiased results with varying guidance. There is a suggestion that a cohort with at least 5% missing data would warrant the use of

multiple imputation but below that there would be negligible benefit[719]. In stark contrast, 5% missingness may be considered the maximum threshold for multiple imputation to be applied[720]. Others have set the minimum threshold to use multiple imputation at 10%[720,721] and 25%[716] as bias is likely at these levels. The maximum threshold to allow for valid results using multiple imputation has been suggested as 40%[720,721].

However, there has been limited studies investigating proportion of missingness in multiple imputations and there is limited evidence to support these thresholds. The few studies that have investigated increasing proportions of missing data for reduction in bias and improvement in efficiency have done so with up to 50% missingness[720,722,723]. However, the results of these have been mixed. For studies investigating greater than 50% missingness, the use of auxiliary variables was not included[724,725]. One study investigated how the percentage of missingness and the incorporation of other variables affected the performance of the multiple imputation model[720]. The top model used included 11 variables in addition to the outcome and exposure variables. They found that if enough variables were used within the multiple imputation model, unbiased results were produced even with up to 90% missingness.

The proportion of missingness in the cohort investigating new surrogate markers of THA survival ranges from 25.1% to 39.8% for each variable with missing data. The total missingness for the cohort was just under 60%. The number of variables included in addition to outcome and exposure variables was 13. Therefore, multiple imputation

is required to ensure unbiased results from complete case analysis and is appropriate with the level of missingness and number of additional variables included.

Distribution of variables

Once the variables to be included are identified, their distribution should be determined. The varying multiple imputation techniques make the assumption that all the data are normally distributed[675]. Using continuous variables that are not normally distributed could result in bias. For example, if body mass index had a skewed distribution the imputation could produce implausible results that are either too low or possibly negative. Categorical or binary variables with missing values is more challenging and certain techniques perform better than others[675]. To handle these variables, such as missing data in smoking status, multiple imputation by chained equations (MICE) is suitable[726]. Furthermore, in the presence of multiple variables of varying types with missing data, MICE has been shown to be a more appropriate method to use[727,728]. Finally, for studies investigating relative risk, MICE has been recommended as the multiple imputation method to use[729]. In contrast to other multiple imputation models that assumes a joint model where two events occur together at the same time point, MICE is a flexible technique to deal with large datasets in which this assumption is rarely valid[728]. Rather than running simple regression models to impute missing values, a series of regression models are performed in which each variable with missing values is modelled conditional to other variables. Instead of a variable modelled with one regression model overall, each variable is modelled according to its own distribution. For example, a continuous variable would be

modelled using a linear regression, a binary variable with logistic regression, and count variables would use a Poisson model[728].

Therefore, all continuous variables used within the final analysis regression model was assessed for normality. Those that were found to be skewed were transformed in an attempt to keep the data continuous. However, if transformation failed to produce a normal distribution, the variable was categorised.

Checking the plausibility of MAR assumption

There are no formal tests to investigate or determine if missing data meet the criteria for missing at random from the data[676]. In addition, the MAR assumption is generally accepted in research using electronic medical records with plausible examples such as women are likely to attend their primary care practice regularly, resulting in a higher chance of their BMI, alcohol or smoking status being recorded[718]. However, it is possible to tabulate each variable used for imputation against the variable with missing data to compare patients with missing data and those without. By doing so, it is possible to evaluate whether it is likely the missing data meet the conditions required to be categorised as MAR[635]. A backward stepwise regression model for registered variables against missingness of imputed variables was performed[718].

Number of imputations

One of the key practical considerations when implementing multiple imputations is to determine the number of imputations (number of datasets) to perform especially when

faced with significant levels of missing data. Initial teachings and studies have argued that only a small number of imputations, such as three to five datasets are required[695,730]. Rubin argued that five imputation datasets is sufficient with a 91% efficiency even when there is up to 50% missing data[699]. Some have suggested the number of imputations is similar to the proportion of missing data[731,732].

However, it is imperative to account for the proportion of missing data and the acceptable deviation from infinite imputations[730]. In the study by Graham et al, statistical power and precision improved with imputed datasets dependant on the proportion of missingness and loss of power tolerance[730]. Table 7-2 shows an adaptation of their findings.

Table 7-2: Number of imputations required to reduce loss of power according to proportion of missingness. Adaptation from Graham et al.

Proportion of Missingness (%)	No. Imputations with Loss of power <5%	No. Imputations with Loss of power <3%	No. Imputations with Loss of power <1%
10	3	5	20
30	10	20	20
50	10	20	40
70	20	40	40
90	40	40	100

As mentioned, the proportion of missingness in the cohort investigating new surrogate markers of THA survival reached just under 60%. Graham et al recommended to use a less than 1% acceptable level of loss of power when statistical power matters. Therefore, for the investigations here 40 imputations were used.

Evaluation of multiple imputation model

For the majority of studies, it is often the case that patients with missing data would differ from those without missing data[635,675]. Therefore, it is important to explore differences between individuals with missing data and those without missing data for key characteristics[733]. Furthermore, comparing the characteristics of patients with missing values against those without missing data can be tabulated to explore the MAR assumption[635].

One method to evaluate whether the multiple imputation has worked appropriately is to compare pre- and post-univariate regression models[675]. In this study, both the pre- and post-multiple imputation Cox proportional regression and Fine & Gray modelling were compared to evaluate the multiple imputation. These should be identical as multiple imputation aims to replicate the original data but with imputed observations and in this cohort, neither the outcome nor the exposure variables had missing data.

Further evaluation of the multiple imputation was performed using Monte Carlo errors during the regression model[734] which is an additional command as part of the multiple imputation package in Stata[698]. White et al suggested using Monte Carlo Errors (MCE) of the multiple imputation result to assess the level of simulation error. The MCE were defined as "the standard deviation of the results across repeated runs of the same imputation procedure using the same data"[734]. There are several aspects the authors proposed to evaluate with the MCE estimates including the parameter estimates, p-

values, confidence intervals and multiple imputation statistics. Guidelines have been proposed by White et al[731]:

1. The MCE estimates of the coefficients should be less than or equal to 10% of the standard errors of the Hazard
2. The MCE estimates of the test-statistic should be less than or equal to 0.1
3. The MCE estimates of the p-value coefficient should be less than or equal to 0.01 when the true p-value is 0.05

7.3. Appendix 3

Survival analysis

Epidemiological studies investigate the occurrence of diseases over time within a population, and the concepts of measuring the frequency of disease occurrence are 'risk' and 'rate'[550]. Risk can be defined as the proportion of patients with the disease from a population originally disease-free over a specified time[735]. In contrast, rate is the term used for the proportion of patients with the disease from the amount of person-time years at risk over a specified time[735]. The risk is also the probability of a patient who is disease free to subsequently develop the disease over a specified time. Therefore, if all patients are observed in a study, the risk can be calculated. However, the vast majority of observational studies inevitably experience loss to follow-up[550].

When investigating associations between exposure and outcome with defined time intervals and where time to outcome of interest is pertinent a particular statistical method is employed, so called 'survival analysis'[547,550]. Investigating survival of hip arthroplasty devices using registry data represents these time intervals: starting from the date of operation and ending with the date of revision surgery[547]. There is an assumption that if followed for a long enough period, all patients will experience revision. Therefore, to calculate the risk of the outcome of interest, all patients undergoing a primary hip replacement must be identified and followed long enough to know their outcome[736]. However, not every patient will undergo revision. Some devices will continue to function well until the end of the follow-up period of the study and therefore not need to be revised. Furthermore, some patients may die before needing revision surgery or can be lost to follow-up due to other reasons. In this

scenario, survival times of a device are incalculable due to incomplete observations and thus data is termed "censored"[547]. In standard analysis, patients who are censored would be excluded and treated in a similar manner to complete case analysis. Furthermore, such analysis assumes normality of the data which is rarely the case[548]. As discussed previously, this could lead to bias as there could be systematic difference between patients who are censored and those included. However, it is a common assumption that censoring occurs randomly[547]. In other words, there must not be any systematically higher or lower risk of the outcome of interest in patients who are censored compared to those who are not[550]. It is agreed that censored data should remain despite not knowing when the outcome of interest occurs, as it is known that the device was not revised prior to censoring having survived up until that point. Therefore, there have been techniques devised to analyse survival in the context of censoring.

These techniques are broadly based on two functions used to describe survival data[548]. The first is the 'survival function' commonly denoted as $S_{(t)}$. $S_{(t)}$ is the probability the device will survive until or beyond a specified time (t)[548,737]. The second is the 'hazard function' commonly denoted as $h_{(t)}$. This is the conditional probability of the outcome of interest (revision in the case of an implant) of occurring given survival to time t[548].

Kaplan-Meier estimates and lifetables

The survival function is commonly represented in either a table or graph and estimated using two different approaches. The first is a lifetable[738] producing the table format and

secondly is the Kaplan-Meier[644] estimate producing the graph. The key to both approaches is the non-parametric nature with no requirement for assumptions to be made regarding the distribution of survival times. They achieve this by estimating the survival function as the cumulative probability that a device survives at each set time interval until time t[547,548]. The lifetables rely on a set of pre-defined intervals to calculate survival intervals. In contrast, the Kaplan-Meier approach reduces potential bias from time intervals chosen subjectively by determining the intervals to include only one outcome of interest event. Therefore, if no censoring occurred the survival function estimated by the Kaplan-Meier method would equal the proportion of devices that are *not* revised at time t_n. Although lifetables can give a useful overview of survival at specific time intervals, the Kaplan-Meier approach is preferred when the timings of revision are known such as in registry data[547]. Lifetables may be biased when faced with heavy censoring, which is often the case near the end of the study period, resulting in the numbers at risk being significantly reduced after the recorded last revision, and confidence intervals subsequently too narrow.

Using lifetables survival can be compared at one time point, but this may not be representative over the study period. Bias can occur at this stage as it is the researcher who chooses the time-point for comparison. Therefore, the Kaplan-Meier method would be more representative of the study period, and test for a null hypothesis of no difference between the survival function of two or more devices. It is possible to compare estimated survival functions from the Kaplan-Meier method, which assess survival function over the whole study period, using the Log-rank test[739]. It is the Kaplan-Meier method that is often used to compare survival rates of different hip

replacement devices[11]. Visually, this can be represented as a survival curve created by plotting $S_{(t)}$ against t[547]. However, there are limitations to using this approach which have been highlighted by recent metal-on-metal failings where single centre cohort studies identified problems before the NJR classed them as outliers[62,65].

The main limitation of the Kaplan-Meier method and log rank test to compare survival functions between devices is the inability to account for other explanatory variables[548]. Therefore, other techniques are required to compare survival times adjusting for confounding variables such as the Cox proportional hazard model ('Cox regression')[547,548].

Cox proportional hazard model

Randomised control trials use randomisation in an attempt to account for unknown confounders. However, observational studies, including joint registry studies, may suffer from potential bias by confounding due to systematic differences between patient groups with the different devices[547]. Therefore, using only the log rank test could affect validity of results and fail to see a true difference. For example, if younger patients who are active were given a particular device, this could have higher revision rates compared to another device given to older patients who are less active.

The Cox regression is a form of multiple regression modelling that investigates survival by estimating the hazard function. The hazard function corresponds to the

concept of rate[550]. In a similar manner to the Kaplan-Meier method, the rate is calculated against the conditional probability of revision before that time given survival. A 'one-to-one' relationship exists between the hazard function and survival function allowing for easy interpretation[550]. No assumptions are made in regard to the probability distribution. However, there is an assumption that the risk between two devices remains proportional over time and thus the same at any given time interval, known as 'proportional hazards assumption'[547]. Therefore, after plotting survival curves, if any cross for the devices investigated, the proportional hazard assumption is violated[547]. It is useful to calculate the hazard rate when conducting survival studies as it can be interpreted as the estimation of the risk of revision per unit time at time t, given survival of the device until time t[550]. Therefore, the hazard rate is said to describe survival 'locally in time' in patients who considered at risk.

Unfortunately, the cox regression model relies on data for potential confounders to be identified and collected prior to modelling. It is possible for residual confounding to remain where even after adjusting for a priori of confounders, a systematic difference is present[547]. However, one significant limitation remains when using Cox regression modelling due to competing risks[631] which can be accounted for by using Fine and Gray proportional hazards model[740].

Fine and Gray proportional hazards model

A 'competing risk' is a variable that would preclude the outcome of interest from occurring[631]. In registry data there are two commonly occurring endpoints recorded:

revision and death[631]. Revision is the outcome of interest in device survival studies, but death is a competing risk as a patient who has died without revision would no longer be at risk of revision surgery. It is possible the presence of a competing risk can bias results from survival analysis such as Kaplan-Meier method[741]. Bias can occur due to the underlying assumption that the risk of revision for either censored or uncensored patients is the same. In the real world setting only what occurs during the lifetime of the patient should be considered. Therefore, the Kaplan-Meier approach is inappropriate to use as it assumes the patient is alive until the revision resulting in bias. If censored patients are precluded from revision, as is the case with death, the risk of revision will be overestimated[631]. In this scenario, the 'cumulative incidence function' can be used to provide an estimation of revision risk that is more valid[742]. A distinction is made with patients who have died considered as competing events and patients who are both alive and without revision at the end of follow-up are considered censored. In Kaplan-Meier method, censored patients are excluded from the at-risk population which would be the same fate for patients who have died[631]. The assumption is patients who are censored have the same risk of revision as patients who are still observed in the study. While this may be valid for patients who are alive and not revised, it is not so for patients who are censored due to death.

Despite the cumulative incidence function addressing the issue of competing risks, it is unable to account for confounding factors. In a scenario where a patient requires revision surgery but does not undergo an operation due to comorbidities, it is vital that comorbidities are included within the model[631]. One study of the Danish Hip Arthroplasty Register showed that with 11.4% of patients dying within a similar time

period, the difference increased between the Kaplan-Meier estimates and cumulative incidence function with increasing follow-up period[631].

Unfortunately, when adjusting for confounders using the Cox regression model, it does not solve the issue of competing risks as results are challenging to interpret[741]. Patients who experience the competing event are not considered within the population at risk of revision. However, cumulative incidence function depends on not only the hazard function as with the Cox regression, but also the 'hazard' of death[550]. Therefore, there no longer exists a one-to-one relationship between the cumulative incidence function and hazard. An alternative approach to address the issue of competing risk is the Fine and Gray (F&G) regression[740]. In this approach patients who have died remain within the at-risk group at each time interval. Furthermore, F&G regression can "directly link the cumulative incidence to explanatory variables" while addressing competing risk interference[550].

There were significantly more patients potentially censored due to mortality compared to loss to follow-up in the RACat; greater than twenty times. Furthermore, the majority of patients (over 70%) in the cohort are over the age of 70. Patients in their mid-seventies at primary hip replacement have a 90% chance of dying before revision is required whereas the converse applies in those under fifty[743]. Death has been shown to be an important competing risk especially in the elderly population[483]. Therefore, a larger proportion of patients who are older within a cohort can compound the effect of competing risk of death in survival analysis[631]. Exposures were also investigated for crude association with mortality rates using Cox regression. In light of these factors,

F&G regression modelling was used in the final analysis to adjust for competing risk of death. The model was applied to each imputed dataset and subsequently combined. However, despite a clear indication to use F&G modelling here, differences between F&G and Cox regression modelling may be small, and it is therefore recommended to always perform both cause-specific Cox and F&G regression models[632].

7.4. Appendix 4

7.4.1. Atypical analgesic multiple imputation results

Baseline characteristics

Table 7-3 shows the availability of the variables with missing data: BMI, Smoking status and Alcohol status. Smoking status had the lowest proportion of missing data with 25.8%, BMI had 34.7% missing and alcohol status had the highest proportion with 41.2%.

Table 7-3: Proportion of missingness for confounding variables with missing data for the Atypical Analgesic cohort.

Characteristic	Missing Observation N (%)
BMI	4,966 (34.7)
Smoking	3,688 (25.8)
Alcohol	5,902 (41.2)

Table 7-4 shows the baseline characteristics of a complete case dataset versus all eligible patients.

Table 7-4: Baseline characteristics of confounding variables with complete case analysis or all eligible patients in the study for the Atypical Analgesic cohort.

Characteristic		Complete Case	All Eligible Patients
BMI	Observed Mean (SD)	28.6 (±4.8)	28.5 (±4.9)
	Missing N (%)	-	4,966 (34.7)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**

Age	40 to <60	544 (9.7)	1,873 (13.1)
	60 to <70	1,061 (18.9)	2,513 (17.6)
	70 to <80	2,057 (36.7)	4,614 (32.2)
	80 and older	1,945 (34.7)	5,308 (37.1)
Gender	Male	2,254 (40.2)	5,702 (39.9)
	Female	3,353 (59.8)	8,606 (60.1)
Alcohol	No	4,043 (72.1)	6,118 (42.8)
	Yes/Ex	1,564 (27.9)	2,288 (16.0)
	Missing	-	5,902 (41.2)
Smoking	Never	4,201 (74.9)	7,918 (55.3)
	Current/Ex	1,406 (25.1)	2,702 (18.9)
	Missing	-	3,688 (25.8)
Charlson Score	0	2,231 (39.8)	6,498 (45.4)
	1	192 (3.4)	687 (4.8)
	≥2	3,184 (56.8)	7,123 (49.8)
Socio-economic status	1	899 (16.0)	2,428 (16.9)
	2	1,004 (17.9)	2,396 (16.8)
	3	985 (17.6)	2,386 (16.7)
	4	1,020 (18.2)	2,398 (16.8)
	5	1,052 (18.8)	2,380 (16.6)
	Rural	647 (11.5)	2,320 (16.2)
CVD	No	5,132 (91.5)	13,296 (92.9)
	Yes	475 (8.5)	1,012 (7.1)
AF	No	5,117 (91.3)	13,261 (92.7)
	Yes	490 (8.7)	1,047 (7.3)
Hypertension	No	1,747 (31.2)	6,507 (45.5)
	Yes	3,860 (68.8)	7,801 (54.5)
Diabetes	No	4,246 (75.7)	11,769 (82.3)
	Yes	1,361 (24.3)	2,539 (17.7)
Depression/Anxiety	No	4,454 (79.4)	11,703 (81.8)
	Yes	1,153 (20.6)	2,605 (18.2)
Cancer	No	5,362 (95.6)	13,691 (95.7)
	Yes	245 (4.4)	617 (4.3)
Indication	Fracture	1,763 (31.4)	4,689 (32.8)
	Osteoarthritis	3,448 (61.5)	8,595 (60.1)
	Other	396 (7.1)	1,024 (7.1)
Previous Fixation	No	5,284 (94.2)	13,646 (95.4)
	Yes	323 (5.8)	662 (4.6)
Previous Fracture	No	4,673 (83.3)	12,191 (85.2)
	Yes	934 (16.7)	2,117 (14.8)

Characteristics following multiple imputation

Baseline characteristics before multiple imputation was compared to post multiple imputation for variables imputed (Table 7-5). All other variables remained the same and will be tabulated later.

Table 7-5: pre-and post-multiple imputation baseline characteristics for BMI, smoking and alcohol variables for the Atypical Analgesic cohort.

Characteristic	Category	Pre-MI	Post-MI
BMI	Observed Mean (SD)	28.5 (±4.9)	28.2 (±4.9)
		Observation N (%)	Observation N (%)
Alcohol	No	6,118 (72.8)	10,402 (72.7)
	Yes/Ex	2,288 (27.2)	3,906 (27.3)
Smoking	Never	7,918 (74.6)	10,688 (74.7)
	Current/Ex	2,702 (25.4)	3,620 (25.3)

Figure 7-1 shows the distribution of BMI, smoking status and alcohol status as imputed in each of the 40 datasets.

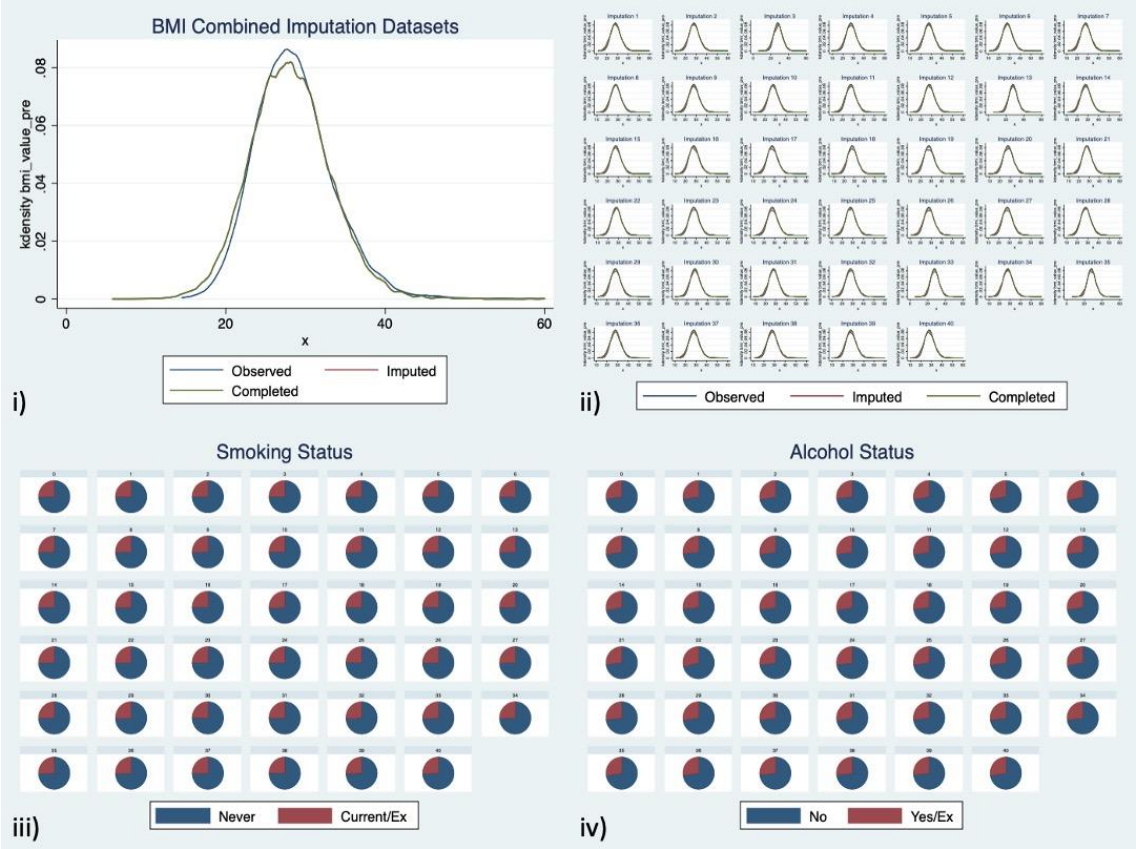


Figure 7-1: Distribution of missing variables following multiple imputation for the Atypical Analgesic cohort. i) BMI histogram within combined datasets. ii) BMI histogram within each dataset. iii) Smoking distribution within each dataset. iv) Alcohol distribution within each dataset.

7.5. Appendix 5

7.5.1. Baseline characteristics for exposures

Table 7-6: Baseline characteristics for the NSAID exposure.

		NSAID 1	NSAID 2	NSAID 3	NSAID 4	NSAID 5
Characteristic						
BMI	Observed Mean (SD)	27.6 (±4.8)	28.2 (±4.7)	28.5 (±4.8)	29.9 (±4.8)	29.8 (±4.9)
	Missing N (%)	1,948 (11.9)	755 (4.6)	1,024 (6.2)	922 (5.6)	860 (5.2)
Characteristic	**Category**	**Observation N (%)**				
Age	40 to <60	642 (3.9)	365 (2.2)	499 (3.0)	342 (2.1)	310 (1.9)
	60 to <70	625 (3.8)	398 (2.4)	614 (3.7)	630 (3.8)	661 (4.0)
	70 to <80	1,190 (7.2)	628 (3.8)	1,111 (6.8)	1,207 (7.3)	1,263 (7.7)
	80 and older	1,998 (12.2)	732 (4.5)	1,058 (6.4)	1,108 (6.7)	1,052 (6.4)
Gender	Male	1,892 (11.5)	1,011 (6.2)	1,376 (8.4)	1,144 (7.0)	1,044 (6.4)
	Female	2,563 (15.6)	1,112 (6.8)	1,906 (11.6)	2,143 (13.0)	2,242 (13.6)
Alcohol	No	1,871 (11.4)	827 (5.0)	1,376 (8.4)	1,582 (9.6)	1,596 (9.7)
	Yes/Ex	673 (4.1)	393 (2.4)	571 (3.5)	531 (3.2)	482 (2.9)
	Missing	1,911 (11.6)	903 (5.5)	1,335 (8.1)	1,174 (7.1)	1,208 (7.4)
Smoking	Never	1,736 (10.6)	1,160 (7.1)	1,940 (11.8)	2,149 (13.1)	2,203 (13.4)
	Current/Ex	628 (3.8)	513 (3.1)	731 (4.4)	626 (3.8)	623 (3.8)
	Missing	2,091 (12.7)	450 (2.7)	611 (3.7)	512 (3.1)	460 (2.8)
Charlson Score	0	1,931 (11.8)	976 (5.9)	1,540 (9.4)	1,466 (8.9)	1,424 (8.7)
	1	220 (1.3)	112 (0.7)	141 (0.9)	147 (0.9)	147 (0.9)
	≥2	2,304 (14.0)	1,035 (6.3)	1,601 (9.7)	1,674 (10.2)	1,715 (10.4)
Socio-economic status	1	834 (5.1)	399 (2.4)	536 (3.3)	521 (3.2)	484 (2.9)
	2	731 (4.4)	382 (2.3)	571 (3.5)	556 (3.4)	533 (3.2)
	3	692 (4.2)	359 (2.2)	591 (3.6)	574 (3.5)	552 (3.4)

	4	661 (4.0)	343 (2.1)	556 (3.4)	592 (3.6)	618 (3.38)
	5	625 (3.8)	316 (1.9)	562 (3.4)	606 (3.7)	661 (4.0)
	Rural	912 (5.5)	324 (2.0)	466 (2.8)	438 (2.7)	438 (2.7)
CVD	No	4,105 (25.0)	1,970 (12.0)	3,040 (18.5)	3,050 (18.6)	3,042 (18.5)
	Yes	350 (2.1)	153 (0.9)	242 (1.5)	237 (1.4)	244 (1.5)
AF	No	4,109 (25.0)	1,951 (11.9)	3,048 (18.5)	3,046 (18.5)	3,084 (18.8)
	Yes	346 (2.1)	172 (1.0)	234 (1.4)	241 (1.5)	202 (1.2)
Hypertension	No	2,222 (13.5)	1,034 (6.3)	1,503 (9.1)	1,330 (8.1)	1,302 (7.9)
	Yes	2,233 (13.6)	1,089 (6.6)	1,779 (10.8)	1,957 (11.9)	1,984 (12.1)
Diabetes	No	3,683 (22.4)	1,756 (10.7)	2,698 (16.4)	2,644 (16.1)	2,661 (16.2)
	Yes	772 (4.7)	367 (2.2)	584 (3.6)	643 (3.9)	625 (3.8)
Depression/Anxiety	No	3,666 (22.3)	1,769 (10.8)	2,718 (16.5)	2,575 (15.7)	2,495 (15.2)
	Yes	789 (4.8)	354 (2.2)	564 (3.4)	712 (4.3)	791 (4.8)
Cancer	No	4,309 (26.2)	2,021 (12.3)	3,141 (19.1)	3,134 (19.1)	3,132 (19.1)
	Yes	146 (0.9)	102 (0.6)	141 (0.9)	153 (0.9)	154 (0.9)
Indication	Fracture	1,608 (9.8)	760 (4.6)	1,047 (6.4)	962 (5.9)	823 (5.0)
	Osteoarthritis	2,579 (15.7)	1,207 (7.3)	2,006 (12.2)	2,076 (12.6)	2,149 (13.1)
	Other	268 (1.6)	156 (0.9)	229 (1.4)	249 (1.5)	314 (1.9)
Previous Fixation	No	4,302 (26.2)	2,032 (12.4)	3,103 (18.9)	3,102 (18.9)	3,105 (18.9)
	Yes	153 (0.9)	91 (0.6)	179 (1.1)	185 (1.1)	181 (1.1)
Previous Fracture	No	3,845 (23.4)	1,812 (11.0)	2,783 (16.9)	2,745 (16.7)	2,760 (16.8)
	Yes	610 (3.7)	311 (1.9)	499 (3.0)	542 (3.3)	526 (3.2)

Table 7-7: Baseline characteristics for the Tramadol exposure.

		Non-User	Tramadol User
Characteristic			
BMI	Observation Mean (SD)	28.4 (±4.8)	29.8 (±4.9)
	Missing N (%)	4,588 (27.9)	921 (5.6)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**
Age	40 to <60	1,767 (10.8)	391 (2.4)
	60 to <70	2,281 (13.9)	647 (3.9)
	70 to <80	4,144 (25.2)	1,255 (7.6)
	80 and older	5,017 (30.5)	931 (5.7)
Gender	Male	5,430 (33.0)	1,037 (6.3)
	Female	7,779 (47.3)	2,187 (13.3)
Alcohol	No	5,776 (35.1)	1,476 (9.0)
	Yes/Ex	2,170 (13.2)	480 (3.0)
	Missing	5,263 (32.0)	1,268 (7.7)
Smoking	Never	7,187 (43.7)	2,001 (12.1)
	Current/Ex	2,487 (15.2)	634 (3.9)
	Missing	3,535 (21.5)	589 (3.6)
Charlson Score	0	5,916 (36.0)	1,421 (8.6)
	1	610 (3.7)	157 (1.0)
	≥2	6,683 (40.7)	1,646 (10.0)
Socio-economic status	1	2,282 (13.9)	492 (3.0)
	2	2,243 (13.6)	530 (3.3)
	3	2,222 (13.5)	546 (3.3)
	4	2,182 (13.3)	588 (3.6)
	5	2,140 (13.0)	630 (3.8)
	Rural	2,140 (13.0)	438 (2.7)
CVD	No	12,249 (74.5)	2,958 (18.0)
	Yes	960 (5.8)	266 (1.6)
AF	No	12,207 (74.3)	3,031 (18.4)
	Yes	1,002 (6.1)	193 (1.2)
Hypertension	No	6,055 (36.9)	1,336 (8.1)
	Yes	7,154 (43.5)	1,888 (11.5)
Diabetes	No	10,821 (65.8)	2,621 (15.9)
	Yes	2,388 (14.6)	603 (3.7)
Depression/Anxiety	No	10,734 (65.3)	2,489 (15.1)
	Yes	2,475 (15.1)	735 (4.5)
Cancer	No	12,681 (77.2)	3,056 (18.6)
	Yes	528 (3.2)	168 (1.0)
Indication	Fracture	4,420 (26.9)	780 (4.7)
	Osteoarthritis	7,883 (48.0)	2,134 (13.0)
	Other	906 (5.5)	310 (1.9)
Previous Fixation	No	12,590 (76.6)	3,054 (18.6)
	Yes	619 (3.8)	170 (1.0)
Previous Fracture	No	11,245 (68.4)	2,700 (16.4)
	Yes	1,964 (12.0)	524 (3.2)

Table 7-8: Baseline characteristics for the Atypical Analgesic exposure.

		Non-User	Atypicals User
Characteristic			
BMI	Observation Mean (SD)	28.5 (±4.9)	29.5 (±4.7)
	Missing N (%)	4,697 (32.8)	269 (1.9)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**
Age	40 to <60	1,777 (12.4)	96 (0.7)
	60 to <70	2,327 (16.3)	186 (1.3)
	70 to <80	4,240 (29.6)	374 (2.6)
	80 and older	5,092 (35.6)	216 (1.5)
Gender	Male	5,412 (37.8)	290 (2.0)
	Female	8,024 (56.1)	582 (4.1)
Alcohol	No	5,770 (40.3)	348 (2.4)
	Yes/Ex	2,168 (15.2)	120 (0.8)
	Missing	5,498 (38.4)	404 (2.8)
Smoking	Never	7,387 (51.6)	531 (3.7)
	Current/Ex	2,534 (17.7)	168 (1.2)
	Missing	3,515 (24.6)	173 (1.2)
Charlson Score	0	6,114 (42.7)	384 (2.7)
	1	634 (4.4)	53 (0.4)
	≥2	6,688 (46.7)	435 (3.0)
Socio-economic status	1	2,295 (16.0)	133 (0.9)
	2	2,244 (15.7)	152 (1.1)
	3	2,238 (15.6)	148 (1.0)
	4	2,233 (15.6)	165 (1.2)
	5	2,230 (15.5)	160 (1.1)
	Rural	2,206 (15.4)	114 (0.8)
CVD	No	12,484 (87.3)	812 (5.7)
	Yes	952 (6.7)	60 (0.4)
AF	No	12,431 (86.9)	830 (5.8)
	Yes	1,005 (7.0)	42 (0.3)
Hypertension	No	6,119 (42.8)	338 (2.7)
	Yes	7,317 (51.1)	484 (3.4)
Diabetes	No	11,074 (77.4)	695 (4.9)
	Yes	2,362 (16.5)	177 (1.2)
Depression/Anxiety	No	11,013 (77.0)	690 (4.8)
	Yes	2,423 (16.9)	182 (1.3)
Cancer	No	12,871 (90.0)	820 (5.7)
	Yes	565 (3.9)	52 (0.4)
Indication	Fracture	4,472 (31.3)	217 (1.5)
	Osteoarthritis	8,011 (56.0)	584 (4.1)
	Other	953 (6.7)	71 (0.5)
Previous Fixation	No	12,818 (89.6)	828 (5.8)
	Yes	618 (4.3)	44 (0.3)
Previous Fracture	No	11,448 (80.0)	743 (5.2)
	Yes	1,988 (13.9)	129 (0.9)

Table 7-9: Baseline characteristics for the frequency of GP visits exposure.

		GP Visits 1-4	GP Visits 0	GP Visits >4
Characteristic				
BMI	Observation Mean (SD)	28.4 (±5.0)	27.7 (±4.8)	28.9 (±4.8)
	Missing N (%)	1,508 (9.2)	1,279 (7.8)	2,722 (16.6)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**	**Observation N (%)**
Age	40 to <60	480 (2.9)	288 (1.8)	1,390 (8.5)
	60 to <70	719 (4.4)	347 (2.1)	1,862 (11.3)
	70 to <80	1,221 (7.4)	677 (4.1)	3,501 (21.3)
	80 and older	1,486 (9.0)	1,255 (7.6)	3,207 (19.5)
Gender	Male	1,582 (9.6)	978 (6.0)	3,907 (23.8)
	Female	2,324 (14.1)	1,589 (9.7)	6,053 (36.8)
Alcohol	No	1,631 (9.9)	1,038 (6.3)	4,583 (27.9)
	Yes/Ex	609 (3.7)	344 (2.1)	1,697 (10.3)
	Missing	1,666 (10.1)	1,185 (7.2)	3,680 (22.4)
Smoking	Never	2,216 (13.5)	598 (3.6)	6,374 (38.8)
	Current/Ex	773 (4.7)	197 (1.2)	2,151 (13.1)
	Missing	917 (5.6)	1,772 (10.8)	1,435 (8.7)
Charlson Score	0	1,890 (11.5)	1,088 (6.6)	4,359 (26.5)
	1	190 (1.2)	144 (0.9)	433 (2.6)
	≥2	1,826 (11.1)	1,335 (8.1)	5,168 (31.4)
Socio-economic status	1	695 (4.2)	485 (3.0)	1,594 (9.7)
	2	667 (4.1)	372 (2.3)	1,734 (10.6)
	3	678 (4.1)	378 (2.3)	1,712 (10.4)
	4	621 (3.8)	372 (2.3)	1,777 (10.8)
	5	589 (3.6)	351 (2.1)	1,830 (11.1)
	Rural	656 (4.0)	609 (3.7)	1,313 (8.0)
CVD	No	3,678 (22.4)	2,367 (14.4)	9,162 5.8)
	Yes	228 (1.4)	200 (1.2)	798 (4.9)
AF	No	3,684 (22.4)	2,364 (14.4)	9,190 (55.9)
	Yes	222 (1.4)	203 (1.2)	770 (4.7)
Hypertension	No	1,887 (11.5)	1,327 (8.1)	4,177 (25.4)
	Yes	2,019 (12.3)	1,240 (7.5)	5,783 (35.2)
Diabetes	No	3,317 (20.2)	2,155 (13.1)	7,970 (48.5)
	Yes	589 (3.6)	412 (2.5)	1,990 (12.1)
Depression/Anxiety	No	3,272 (19.9)	2,074 (12.6)	7,877 (47.9)
	Yes	634 (3.9)	493 (3.0)	2,083 (12.7)
Cancer	No	3,752 (22.8)	2,505 (15.2)	9,480 (57.7)
	Yes	154 (0.9)	62 (0.4)	480 (2.9)
Indication	Fracture	1,441 (8.8)	800 (4.9)	2,959 (18.0)
	Osteoarthritis	2,190 (13.3)	1,604 (9.8)	6,223 (37.9)
	Other	275 (1.7)	163 (1.0)	778 (4.7)
Previous Fixation	No	3,712 (22.6)	2,512 (15.3)	9,420 (57.3)
	Yes	194 (1.2)	55 (0.3)	540 (3.3)
Previous Fracture	No	3,329 (20.3)	2,303 (14.0)	8,313 (50.6)
	Yes	577 (3.5)	264 (1.6)	1,647 (10.0)

Table 7-10: Baseline characteristics for the referral to the Orthopaedic department exposure.

		Non-Referral	Ortho Referral
Characteristic			
BMI	Observation Mean (SD)	28.6 (±4.9)	29.2 (±4.9)
	Missing N (%)	5,216 (31.7)	293 (1.8)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**
Age	40 to <60	1,992 (12.1)	166 (1.0)
	60 to <70	2,697 (16.4)	231 (1.4)
	70 to <80	4,999 (30.4)	400 (2.4)
	80 and older	5,605 (34.1)	343 (2.1)
Gender	Male	6,049 (36.8)	418 (2.5)
	Female	9,244 (56.3)	722 (4.4)
Alcohol	No	6,651 (40.5)	601 (3.7)
	Yes/Ex	2,434 (14.8)	216 (1.3)
	Missing	6,208 (37.8)	323 (2.0)
Smoking	Never	8,419 (51.2)	769 (4.7)
	Current/Ex	2,889 (17.6)	232 (1.4)
	Missing	3,985 (24.2)	139 (0.8)
Charlson Score	0	6,831 (41.6)	506 (3.1)
	1	722 (4.4)	45 (0.3)
	≥2	7,740 (47.1)	589 (3.6)
Socio-economic status	1	2,585 (15.7)	189 (1.2)
	2	2,571 (15.6)	202 (1.2)
	3	2,576 (15.7)	192 (1.2)
	4	2,559 (15.6)	211 (1.3)
	5	2,559 (15.6)	211 (1.3)
	Rural	2,443 (14.9)	135 (0.8)
CVD	No	14,169 (86.2)	1,038 (6.3)
	Yes	1,124 (6.8)	102 (0.6)
AF	No	14,179 (86.3)	1,059 (6.4)
	Yes	1,114 (6.8)	81 (0.5)
Hypertension	No	6,890 (41.9)	501 (3.0)
	Yes	8,403 (51.1)	639 (3.9)
Diabetes	No	12,525 (76.2)	917 (5.6)
	Yes	2,768 (16.8)	223 (1.4)
Depression/Anxiety	No	12,367 (75.3)	856 (5.2)
	Yes	2,926 (17.8)	284 (1.7)
Cancer	No	14,640 (89.1)	1,097 (6.7)
	Yes	653 (4.0)	43 (0.3)
Indication	Fracture	4,881 (29.7)	319 (1.9)
	Osteoarthritis	9,272 (56.4)	745 (4.5)
	Other	1,140 (6.9)	76 (0.5)
Previous Fixation	No	14,588 (88.8)	1,056 (6.4)
	Yes	705 (4.3)	84 (0.5)
Previous Fracture	No	13,029 (79.3)	916 (5.6)
	Yes	2,264 (13.8)	224 (1.4)

Table 7-11: Baseline characteristics for the referral to the Physiotherapy department exposure.

		Non-Referral	Physio Referral
Characteristic			
BMI	Observation Mean (SD)	28.7 (±4.9)	28.7 (±4.8)
	Missing N (%)	5,183 (31.5)	326 (2.0)
Characteristic	**Category**	**Observation N (%)**	**Observation N (%)**
Age	40 to <60	1,972 (12.0)	186 (1.1)
	60 to <70	2,725 (16.6)	203 (1.2)
	70 to <80	4,940 (30.1)	459 (2.8)
	80 and older	5,473 (33.3)	475 (2.9)
Gender	Male	6,050 (36.8)	417 (2.5)
	Female	9,060 (55.1)	906 (5.5)
Alcohol	No	6,508 (39.6)	744 (4.5)
	Yes/Ex	2,434 (14.8)	216 (1.3)
	Missing	6,168 (37.5)	363 (2.2)
Smoking	Never	8,298 (50.5)	890 (5.4)
	Current/Ex	2,857 (17.4)	264 (1.6)
	Missing	3,955 (24.1)	169 (1.0)
Charlson Score	0	6,815 (41.5)	522 (3.2)
	1	710 (4.3)	57 (0.3)
	≥2	7,585 (46.2)	744 (4.5)
Socio-economic status	1	2,503 (15.2)	271 (1.6)
	2	2,554 (15.5)	219 (1.3)
	3	2,557 (15.6)	211 (1.3)
	4	2,523 (15.4)	247 (1.5)
	5	2,522 (15.3)	248 (1.5)
	Rural	2,451 (14.9)	127 (0.8)
CVD	No	14,017 (85.3)	1,190 (7.2)
	Yes	1,093 (6.7)	133 (0.8)
AF	No	14,038 (85.4)	1,200 (7.3)
	Yes	1,072 (6.5)	123 (0.7)
Hypertension	No	6,857 (41.7)	534 (3.2)
	Yes	8,253 (50.2)	789 (4.8)
Diabetes	No	12,399 (75.5)	1,043 (6.3)
	Yes	2,711 (16.5)	280 (1.7)
Depression/Anxiety	No	12,225 (74.4)	998 (6.1)
	Yes	2,885 (17.6)	325 (2.0)
Cancer	No	14,461 (88.0)	1,276 (7.8)
	Yes	649 (3.9)	47 (0.3)
Indication	Fracture	4,757 (28.9)	443 (2.7)
	Osteoarthritis	9,258 (56.3)	759 (4.6)
	Other	1,095 (6.7)	121 (0.7)
Previous Fixation	No	14,388 (87.6)	1,256 (7.6)
	Yes	722 (4.4)	67 (0.4)
Previous Fracture	No	12,922 (78.6)	1,023 (6.2)
	Yes	2,188 (13.3)	300 (1.8)

7.6. Appendix 6

NSAID use

Identical univariate Cox and F&G modelling was found comparing pre- and post-multiple imputation regressions respectively (Table 7-12).

Table 7-12: Comparison of pre- and post-multiple imputation univariate regression for the NSAID exposure in both Cox proportional hazard and Fine & Gray models.

		PRE MI (95% CI)	POST MI (95% CI)
Cox Proportional Regression	Second Quintile compared to First Quintile (p<20)	0.69 (0.36-1.33) p=0.270	0.69 (0.36-1.33) p=0.270
	Third Quintile compared to First Quintile (p<20)	0.84 (0.48-1.45) p=0.525	0.84 (0.48-1.45) p=0.525
	Fourth Quintile compared to First Quintile (p<20)	1.07 (0.64-1.80) p=0.788	1.07 (0.64-1.80) p=0.788
	Fifth Quintile (p≥80) compared to First Quintile (p<20)	1.65 (1.02-2.65) p=0.041	1.65 (1.02-2.65) p=0.041
Fine & Gray Regression	Second Quintile compared to First Quintile (p<20)	0.80 (0.41-1.55) p=0.511	0.80 (0.41-1.55) p=0.511
	Third Quintile compared to First Quintile (p<20)	1.00 (0.58-1.73) p=0.997	1.00 (0.58-1.73) p=0.997
	Fourth Quintile compared to First Quintile (p<20)	1.32 (0.79-2.21) p=0.297	1.32 (0.79-2.21) p=0.297
	Fifth Quintile (p≥80) compared to First Quintile (p<20)	2.03 (1.26-3.27) **p=0.003**	2.03 (1.26-3.27) **p=0.003**

The results based on 40 imputations satisfy the practical guidelines from White et al[734] and therefore can be reasonably certain about the statistical reproducibility of the results of this study. The Monte Carlo errors are shown in Table 7-13.

Table 7-13: Multiple imputation Monte Carlo errors for the NSAID exposure.

	Coefficient	T-Statistic	P-Value
Second Quintile compared to First Quintile (p<20)	0.00076	0.00	0.001
Third Quintile compared to First Quintile (p<20)	0.00088	0.00	0.003
Fourth Quintile compared to First Quintile (p<20)	0.00107	0.00	0.001
Fifth Quintile (p≥80) compared to First Quintile (p<20)	0.00185	0.01	0.000

Pre- and post-multiple imputation of the multivariate F&G analysis is found in Table 7-14.

Table 7-14: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the NSAID exposure.

		PRE MI (95% CI)	POST MI (95% CI)
Fine & Gray Regression	Second Quintile compared to First Quintile (p<20)	5.72 (1.51-1584.43) p=0.511	0.78 (0.40-1.51) p=0.458
	Third Quintile compared to First Quintile (p<20)	6.30 (1.64-961.63) p=0.997	1.00 (0.58-1.73) p=0.995
	Fourth Quintile compared to First Quintile (p<20)	10.21 (1.91-4086.56) p=0.297	1.41 (0.83-2.39) p=0.206
	Fifth Quintile (p≥80) compared to First Quintile (p<20)	46.63 (2.95-859082.30) **p=0.003**	2.20 (1.35-3.59) **p=0.002**

Tramadol use

Identical univariate Cox and F&G modelling was found comparing pre- and post-multiple imputation regressions respectively (Table 7-15).

Table 7-15: Comparison of pre- and post-multiple imputation univariate regression for the Tramadol exposure in both Cox proportional hazard and Fine & Gray models.

		PRE MI (95% CI)	POST MI (95% CI)
Cox Proportional Regression	Tramadol Users vs. Non-Users	2.16 (1.58-2.97) **p<0.001**	2.16 (1.58-2.97) **p<0.001**
Fine & Gray Regression	Tramadol Users vs. Non-Users	2.36 (1.72-3.23) **p<0.001**	2.36 (1.72-3.23) **p<0.001**

The results based on 40 imputations satisfy the practical guidelines from White et al[734] and therefore can be reasonably certain about the statistical reproducibility of the results of this study. The Monte Carlo errors are shown in Table 7-16.

Table 7-16: Multiple imputation Monte Carlo errors for the Tramadol exposure.

	Coefficient	T-Statistic	P-Value
Tramadol Users vs. Non-Users	0.00110	0.01	0.000

Pre- and post-multiple imputation of the multivariate F&G analysis is found in Table 7-17.

Table 7-17: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for Tramadol exposure.

		PRE MI (95% CI)	POST MI (95% CI)
Fine & Gray Regression	Tramadol Users vs. Non-Users	7.71 (3.15-37.88)	2.42 (1.74-3.35)
		p=0.015	**p<0.001**

Atypical analgesia

Identical univariate Cox and F&G modelling was found comparing pre- and post-multiple imputation regressions respectively (Table 7-18).

Table 7-18: Comparison of pre- and post-multiple imputation univariate regression for the Atypical Analgesic exposure in both Cox proportional hazard and Fine & Gray models.

		PRE MI (95% CI)	POST MI (95% CI)
Cox Proportional Regression	Atypicals Users vs. Non-Users	2.51 (1.67-3.78)	2.51 (1.67-3.78)
		p<0.001	**p<0.001**
Fine & Gray Regression	Atypicals Users vs. Non-Users	2.76 (1.84-4.13)	2.76 (1.84-4.13)
		p<0.001	**p<0.001**

The results based on 40 imputations satisfy the practical guidelines from White et al[734] and therefore can be reasonably certain about the statistical reproducibility of the results of this study. The Monte Carlo errors are shown in Table 7-19.

Table 7-19: Multiple imputation Monte Carlo errors for the Atypical Analgesic exposure.

	Coefficient	T-Statistic	P-Value
Atypicals Users vs. Non-Users	0.0013551	0.01	0.000

Pre- and post-multiple imputation of the multivariate F&G analysis is found in Table 7-20.

Table 7-20: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the Atypical Analgesic exposure.

		PRE MI (95% CI)	POST MI (95% CI)
Fine & Gray Regression	Atypicals Users vs. Non-Users	2.17 (1.51-4.30)	2.59 (1.70-3.95)
		p<0.001	**p<0.001**

GP visits

Identical univariate Cox and Fine and Gray modelling was found comparing pre- and post-multiple imputation regressions respectively (Table 7-21).

Table 7-21: Comparison of pre- and post-multiple imputation univariate regression for the frequency of GP visits exposure in both Cox proportional hazard and Fine & Gray models.

		PRE MI (95% CI)	**POST MI (95% CI)**
Cox Proportional Regression	GP Visits 0 vs. GP Visits 1-4	2.45 (1.23-4.88) **p=0.011**	2.45 (1.23-4.88) **p=0.011**
	GP Visits >4 vs. GP Visits 1-4	1.92 (1.18-3.10) **p=0.008**	1.92 (1.18-3.10) **p=0.008**
Fine & Gray Regression	GP Visits 0 vs. GP Visits 1-4	1.98 (1.00-3.95) p=0.052	1.98 (1.00-3.95) p=0.052
	GP Visits >4 vs. GP Visits 1-4	2.12 (1.31-3.43) **p=0.002**	2.12 (1.31-3.43) **p=0.002**

The results based on 40 imputations satisfy the practical guidelines from White et al[734] and therefore can be reasonably certain about the statistical reproducibility of the results of this study. The Monte Carlo errors are shown in Table 7-22.

Table 7-22: Multiple imputation Monte Carlo errors for the frequency of GP visits exposure.

	Coefficient	T-Statistic	P-Value
GP Visits 0 vs. GP Visits 1-4	0.0011761	0.00	0.000
GP Visits >4 vs. GP Visits 1-4	0.0006818	0.00	0.000

Pre- and post-multiple imputation of the multivariate F&G analysis is found in Table 7-23.

Table 7-23: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the frequency of GP visits exposure.

		PRE MI (95% CI)	POST MI (95% CI)
Fine & Gray Regression	GP Visits 0 vs. GP Visits 1-4	6.04 (1.20-61337379.82) p=0.617	2.07 (1.02-4.18) **p=0.043**
	GP Visits >4 vs. GP Visits 1-4	13.05 (2.54-1185.53) p=0.068	2.10 (1.29-3.42) **p=0.003**

Orthopaedic referral

Identical univariate Cox and Fine and Gray modelling was found comparing pre- and post-multiple imputation regressions respectively (Table 7-24).

Table 7-24: Comparison of pre- and post-multiple imputation univariate regression for the referral to the Orthopaedic department exposure in both Cox proportional hazard and Fine & Gray models.

		PRE MI (95% CI)	POST MI (95% CI)
Cox Proportional Regression	Non-Referral vs. Ortho Referral	1.84 (1.12-3.02) **p=0.015**	1.84 (1.12-3.02) **p=0.015**
Fine & Gray Regression	Non-Referral vs. Ortho Referral	1.95 (1.19-3.20) **p=0.008**	1.95 (1.19-3.20) **p=0.008**

The results based on 40 imputations satisfy the practical guidelines from White et al[734] and therefore can be reasonably certain about the statistical reproducibility of the results of this study. The Monte Carlo errors are shown in Table 7-25.

Table 7-25: Multiple imputation Monte Carlo errors for the referral to the Orthopaedic department exposure.

	Coefficient	T-Statistic	P-Value
Non-Referral vs. Ortho Referral	0.0009506	0.00	0.000

Pre- and post-multiple imputation of the multivariate F&G analysis is found in Table 7-26.

Table 7-26: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the referral to the Orthopaedic department exposure.

		PRE MI (95% CI)	POST MI (95% CI)
Fine & Gray Regression	Non-Referral vs. Ortho Referral	15.08 (3.91-220.80) **p=0.004**	1.90 (1.16-3.13) **p=0.011**

Physiotherapy department

Identical univariate Cox and Fine and Gray modelling was found comparing pre- and post-multiple imputation regressions respectively (Table 7-27).

Table 7-27: Comparison of pre- and post-multiple imputation univariate regression for the referral to the Physiotherapy department exposure in both Cox proportional hazard and Fine & Gray models.

		PRE MI (95% CI)	POST MI (95% CI)
Cox Proportional Regression	Non-Referral vs. Physio Referral	2.00 (1.26-3.17) **p=0.003**	2.00 (1.26-3.17) **p=0.003**
Fine & Gray Regression	Non-Referral vs. Physio Referral	2.06 (1.30-3.27) **p=0.002**	2.06 (1.30-3.27) **p=0.002**

The results based on 40 imputations satisfy the practical guidelines from White et al[734] and therefore can be reasonably certain about the statistical reproducibility of the results of this study. The Monte Carlo errors are shown in Table 7-28.

Table 7-28: Multiple imputation Monte Carlo errors for the referral to the Physiotherapy department exposure.

	Coefficient	T-Statistic	P-Value
Non-Referral vs. Ortho Referral	0.0005998	0.00	0.000

Pre- and post-multiple imputation of the multivariate F&G analysis is found in Table 7-29.

Table 7-29: Comparison of pre- and post-multiple imputation multivariate Fine & Gray regression for the referral to the Physiotherapy department exposure.

		PRE MI (95% CI)	POST MI (95% CI)
Fine & Gray Regression	Non-Referral vs. Physio Referral	10.20 (3.22-101.07) **p=0.016**	2.07 (1.29-3.32) **p=0.003**