Predictive Factors and Outcomes in Primary Hip Arthroplasty

M. Abdulhadi Alagha

Imperial College London
Department of Surgery and Cancer
MSk Lab

A thesis submitted to

Imperial College London

in partial fulfilment of

the requirements

for the degree of

Doctor of Philosophy

September, 2023

Originality Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where stated otherwise in the acknowledgement section and that this work has not been submitted for any other degree or professional qualification.

Copyright Declaration

Abstract

Predictive Factors and Outcomes in Primary Hip Arthroplasty
Doctor of Philosophy in Clinical Medicine Research
M A Alagha, MSk Lab, September 2023

Cognisant prognostication involves informed predictions and is of a particular importance in surgical practice where significant variations prevail for patients with end stage hip pathologies necessitating arthroplasty. A comprehensive evaluation of Machine Learning (ML) models was undertaken using two of the largest hip registries, from Sweden and the United Kingdom. Failure rates (re-operation, revision, periprosthetic femoral fracture, and mortality), and patient-reported outcome measures (PROMs) were forecasted for primary elective total hip replacement. ML algorithms had a superior binary discriminative power in predicting all outcomes and achieved "good" to "excellent" accuracy in forecasting significant and substantial clinical improvement for PROMs.

Multivariate analysis of the LASSO*-penalised CoxNet** survival model revealed a statistically significant trade-off encompassing a heightened mortality risk in patients with low body mass index and an augmented revision risk in obese individuals. Likewise, whilst cemented fixation implants were associated with a diminished likelihood of re-do procedures, they were observed to have an increased mortality risk. SHAR data was further evaluated to a) predict the risk of femoral fractures requiring re-do surgery, and b) identify the effect of fixation methods on outcomes in a matched cohort. Cementless fixation, appeared to increase this likelihood with statistically significant differences between the matched cohorts.

Two laboratory-based experiments investigated the effect of impaction energy levels on bone strain behaviour, peri-prosthetic femoral fracture risk, and implant stability using two types of specimens (synthetic and cadaveric hips) and two biomechanical approaches (digital image correlation and rosette strain gauges). The findings of these seem to suggest that a high number of low-energy strikes during femoral broaching and implant seating reduces the risk of femoral fractures without affecting implant fixation.

Using personalised risk stratification models in orthopaedic practice may help balance safety and functional outcomes within the context of variability in patient characteristics, surgical practices, and implant designs.

* Least Absolute Shrinkage and Selection Operator (LASSO)
** Cox proportional hazard regression

Acknowledgements

I must first thank my principal supervisors, Professor Justin Cobb, and Professor Jonathan Jeffers, who provided unwavering support, patience, and guidance throughout these studies. I want to express my deep gratitude to Professor Cobb for being an inspiring role model and mentor. His guidance and support have profoundly influenced my personal and academic journey, and I owe him a deep debt of gratitude. Professor Jeffers opened my mind to the mechanical engineering field and showed me insights for which I am very grateful. My thanks also to Mr Alexander Liddle for his supervision, and for his encouragement and belief in my ability to complete this work. Mr Liddle's doctoral thesis influenced the structure of this PhD thesis.

I have relied on Professor Ola Rolfson, Dr Maziar Mohaddes, Dr Kostas Kalogeropoulos and Dr Ruben Doyle at various stages and without their kind guidance this work could not have been completed. Dr Fabio de Mello provided valuable help in running our algorithms on the NJR dataset. Thanks to Dr Leo Celi for his mentorship and insights on health informatics.

I am immensely thankful for the support of my colleagues within and outside the MSk Lab and Biomechanics Lab. Of particular, Dr Richard Abel, Ms Robyn Andrews, Dr Susannah Clarke, Dr Camilla Halewood, Dr Simon Harris, Ms Kathy Lewis, Dr Geoffrey Ng, and Ms Jessica Prestt, Ms Jennifer Simeon, and Ms Eleanor Westhead, who have always been happy to devote their time and support me when needed. I would also like to acknowledge Dr Mikko Venäläinen who assisted me when needed in troubleshooting parts of the code in Chapters 3 and 4 related to the Swedish Hip Arthroplasty Register. Thanks also to two previous BSc students, Chloe Jordan and Rishabh Gupta, who undertook their intercalated BSc degrees under my co-supervision through the work presented in Chapter 8 of this thesis, and Dr Shabnam Samsami who assisted me in the rosette strain gauges preparation and the testing of four specimens in the study presented in Chapter 9. This work was generously funded by Imperial College London through their President's PhD scholarship.

A special thanks to my friends, Alain-Frederic Guilpain, Matthew Riddell, and Shahid Sharif for their unconditional support and encouragement throughout this journey. I also thank the friends I have made at Imperial for their comradeship. Finally, I thank my siblings for their unconditional love and for paving the way for me to start and complete this work.

To my parents

Contents

Figures

Tables

Chapter 1 Introduction and literature review

1.1. Background and aims

Considerable variability exists in surgical practice and implant selection for patients requiring elective hip arthroplasty. A multifaceted interplay of various factors influences the choice of implants and surgical approaches. The complex and intricate variations amongst patients, surgeons and units make it difficult to detect and compare the relative outcomes for different patients (1). Data from national joint registries, despite their observational nature and susceptibility to biases, continues to form a valuable source for assessing safety and implant survival outcomes in arthroplasty (2-4). However, the significant variation in reported outcomes across registries (2, 5-7), coupled with the limitations of conventional statistical models and their linearity assumption (8-10), prompt the need for further research into possible patterns and relationships and to mitigate some of the confounding effects such as selection bias and confounding by indication (11).

Traditional statistical models have a limited ability to handle the intricacies of complex datasets and are thus unable to address the unique characteristics of individual patients more thoroughly. Modifiable factors by either patient or implant type selection and surgical practice provides an important avenue to improve outcomes following primary elective hip arthroplasty. For instance, an observational survival analysis study using the National Joint Registry of England and Wales (NJR) between 2003 and 2011 for primary elective total hip arthroplasty showed that the type of prosthesis is not related to a significant change in mortality (12). However, a contrasting study using propensity score matching techniques to minimise confounding by indication showed a protective risk for hip resurfacing arthroplasty in comparison to cemented and cementless total hip

arthroplasty (THA) (13). Some of the observed protective exposures in registry data analyses, such as the use of hip resurfacing arthroplasty (HRA) or unicompartemental knee replacement (UKA), are reported to be difficult to interpret and arguably attributed to a combination of patient selection factors and casual effects (14). This discrepancy indicates that the variations in predictor factors used for inference testing in conventional statistical models lead to diverse rates in forecasting and assessing outcomes of interest.

Machine learning techniques have the capability to address some of these challenges through leveraging advanced predictive models and processing complex data patterns whilst minimising confounding factors related to selection bias. The integration of the fields of computing, statistics and hip replacement holds promise in enhancing the predictive accuracy of outcomes and have the potential to empower patients and surgeons in improving the shared decision-making process.

1.2. Osteoarthritis and the hip

Osteoarthritis (OA) is a degenerative joint condition affecting more than half a billion people worldwide and is a leading cause for disability amongst older adults (15). OA poses a significant public health challenge, and its 7% global prevalence is thought to be underestimated due to variations in definitions, sampled populations and diagnostic criteria (16-18).

In the hip, the point prevalence of radiographic hip OA in adults doubled between 1991 and 2018 (28% and 53%, respectively), whilst symptomatic hip OA remained around one-third (36% and 30%, respectively) with higher figures for adults over the age of 75 (19, 20). Several studies have investigated the prevalence of early hip OA using more sensitive Magnetic Resonance Imaging (MRI) in asymptomatic individuals. They revealed the

presence of labral tears in 69% of the joints (21), commonly associated with hip OA (22), whilst high impact athletes exhibited findings consistent with early hip OA irrespective of the presence of symptoms in over 50% of the screened hips (23).

At a cellular level, chondrocytes play a vital role in preserving cartilage structure and orchestrate an equilibrium between anabolic and catabolic cellular signals. The harmony is disrupted in OA pathogenesis favouring higher catabolic activity. This results in subsequent inflammation, cartilage degradation and subchondral bone remodelling, which is evident in radiographs by the appearances of osteophytes formation, subchondral sclerosis and cysts (24). Oscillating levels of pain, stiffness and reduced mobility constitute the predominant symptoms of hip osteoarthritis and depends on several of intrinsic and extrinsic variables. These risk factors include increasing age, female gender, high body mass index (BMI), family history, previous hip injuries, congenital hip abnormality, joint malalignment, low and high intensity levels of physical activity, overall health and diurnal variations (25, 26). A complex and enigmatic correlation exists between hip symptoms and radiographic indicators of disease progression with a proportion of individuals who remain pain-free despite severe radiological OA findings (27-29).

That said, end-stage osteoarthritis requiring hip arthroplasty is on the rise. In the United Kingdom, the number of primary total hip replacements (THA) increased by 3.35% between 2018 (n = 94,936) and 2020 (n = 98,114) and is projected to further increase by 37.7% by the year 2060 (30). In the USA, the total annual count for THA increased from 370,770 in 2014 to projected figures of 498,000 in 2020 (exact number not reported) and 1,429,000 by the year 2040, indicating a rise from 2014 of approximately 34% and 284%, respectively (31). The global burden of hip OA, as illustrated by these figures, underscores the increasing need for strategies aimed at its prevention and management.

1.3. Non-arthroplasty treatment for hip osteoarthritis

Referral criteria for hip replacement in the United Kingdom heavily relies on the Oxford Hip Scores (OHS), and are set inappropriately low favouring cost over clinical-effectiveness (32, 33). The National Institute of Clinical Excellence (NICE) recommends the use of clinical assessments rather than severity scoring systems prior to patient referral to specialist orthopaedics services (34). Amongst patients considered for hip arthroplasty by a specialist surgeon, one-third to one-half are considered ineligible primarily due to not meeting the OHS and pain thresholds indicated for hip replacement (35, 36).

Most hip osteoarthritis treatments aim to manage symptoms and slow disease progression whilst hip arthroplasty is considered curative and reserved for end-stage disease (37). Latest evidence-based guidelines suggest arthritis education and structured land-based therapeutic exercise programmes as the core non-surgical treatment for hip OA, coupled with strong conditional recommendations for mind-body exercise (such as Tai Chi or Yoga), weight management (if BMI ≥30 kg/m2), gait aids and/or non-steroidal anti-inflammatory drugs (NSAIDs) (38, 39). There is only a weak evidence for the use of other non-pharmacological therapies, such routine braces and orthotics, acupuncture, electrotherapy, and manual therapy and thus these are not recommended by NICE (34).

With respect to pharmacological management analgesics, namely NSAIDs should be offered only if needed and alongside non-pharmacological therapies, starting with topical then oral if ineffective, whilst avoiding acetaminophen and weak opioids due to their limited to no benefit (34, 38). Glucosamine and/or chondroitin, commonly used for knee OA, have weak and inconsistent evidence to support their use to alleviate osteoarthritic pain and are not endorsed for usage in the UK (34, 40-42).

Intra-articular hip injections have witnessed a notable upsurge in their application for hip osteoarthritis. Corticosteroid (CS) injections are recommended by NICE and the American Academy of Orthopaedic Surgeons (AAOS) when other non-surgical treatment modalities fail (34, 43). CS appears to variably ameliorate OA symptoms for up to ten weeks, however almost 50% of patients were reported to require definitive surgical treatment within two years (44-46). In contrast, injected hyaluronic acid exhibits a slower onset of action compared to CS, and its effect on hip OA symptoms is less prominent, leading to its non-recommendation by neither NICE nor AAOS (34, 43, 47, 48). The emerging field of cartilage regeneration presents substances that have garnered significant interest such as platelet-rich plasma (49, 50) and mesenchymal stem cells injections (51, 52). However, there is currently limited high-quality evidence to support their use for symptomatic hip osteoarthritis (53). Other promising novel injections that are being investigated include bone marrow aspirate concentrate (BMAC) (54), NSAIDs (55, 56), and prolotherapy (57).

Hip preservation procedures and hip arthroscopy, such as osteochondroplasty and labral repair/debridement, lack substantial evidence of effectiveness and are not recommended by NICE for the management of hip OA (34). The evidence for hip arthroscopic repair for patients with early femoroacetabular impingement (FAI) or chondro-labral lesions is contradictory and mainly consists of case series (level 4 evidence) and methodological limitations reporting varying levels of efficacy (58, 59). A meta-analysis of three RCTs of three short-term patient-reported outcomes comparing hip arthroscopy to physical therapy (PT) for symptomatic FAI showed superior outcomes favouring arthroscopy (60). Nevertheless, several limitations hinder the generalisability of findings, some of which are: the lack of blinding and availability of the minimal clinically important difference (MCID) across the three studies, selection bias of individuals to participate, high crossover between treatment groups and the error in treatment effect that was reported in one study.

Another RCT comparing arthroscopic acetabular labral repair to PT in patients over 40 years of age reported superior outcomes at 12 months, however, their study had several limitations including the lack of statistical power, high crossover to the PT treatment arm, selection bias, and the loss to follow-up (61). A previous systematic review of eleven studies reported a mean time conversion between hip arthroscopy and total hip replacement to be 13.5 months (62), highlighting that hip arthroplasty remains the procedure of choice for patients with symptomatic end-stage hip osteoarthritis.

1.4. Hip replacement

Initial attempts at developing hip arthroplasty were started in Germany by Themistocles Glück in 1891 who used ivory to replace osteonecrosis of the femoral heads that had been caused by tuberculosis (63). Several failed attempts followed in the first half of 20th century using glass and stainless steel materials (64, 65). The first metal-on-metal (MoM) prosthesis was developed by English surgeon George McKee in 1953 using the cemented Thompson femoral stems and cobalt-chromium acetabular cup, but this method was abandoned in the 1970s due to metal particle-related complications (66, 67).

During the 1960s, Sir John Charnley devised the concept of low friction arthroplasty, the cornerstone of modern total hip arthroplasty (THA), consisting of three elements: acrylic bone cement, a metal femoral stem and a polyethylene acetabular component (68). Charnley's subsequent developments to his technique included the use of a smaller femoral head size to reduce the risk of dislocation and increase the range of motion; the use of ultrahigh molecular weight polyethylene (UHMWPE) bearing surfaces to minimise the risks of osteolysis and aseptic loosening; and the introduction of poly-methyl-methacrylate (PMMA) bone cement to improve fixation (66, 69). Overall, Charnley's low-friction THA showed survivorship of 77-86.5% at 25 years follow-up (70).

Significant advancements to implant design, materials and bearing surfaces were developed in the field of hip arthroplasty following Charnley's principles. Trends shifting fixation type from cemented to uncemented started in the 1970s in response to concerns about "cement disease" (71), the introduction of ceramic-on-ceramic bearings in the 1980s, and the refinement of ultra-high cross-linked polyethylene (UHXLPE) and other advanced materials such as BIOLOX forte and delta implants during the 1990s (66). Fixation type in primary hip arthroplasty is considered in more detail in Section 1.10.

On the other hand, hip resurfacing arthroplasty evolved in concurrence with development of THR, building upon the pioneering work of Marius Smith-Petersen and the principle of interposing a perishable barrier (64). Later, Gerard (72) and Michael Freeman (73) developed the cemented MoM and metal-on-poly (MoP) techniques, while the Corin group and Derek McMinn laid the groundwork for modern hip resurfacing arthroplasty in the 1990s (66, 74). Novel implant advancements such as ceramic hip resurfacing arthroplasty (75) are currently being evaluated. Analysis of the rationale, development and results from total hip replacement and hip resurfacing arthroplasty are described in Sections 1.5 and 1.6.

1.5. Total Hip Replacement

1.5.1. Indications

Whilst total hip arthroplasty was initially developed as a measure to treat trauma and tubercular arthritis, primary hip osteoarthritis is the leading cause for THR nowadays. When Sir John Charnley performed the first successful elective hip replacement in 1962 it was to alleviate hip OA symptoms and was considered a procedure for the 'elderly'. As highlighted in previous sections, significant advancements in implant design and

techniques coupled with a better understanding of complications and implant survival rates led to the increasing incidences of THR worldwide for different patient groups. The principal indications for total hip replacement include end-stage osteoarthritis and idiopathic necrosis of the femoral head (AVN). Posttraumatic arthritis, metastatic tumours, congenital hip abnormalities, and inflammatory arthritis rheumatoid arthritis (RA) as well as ankylosing spondylitis (AS) are also indications for THR, albeit less common (76). THR due to trauma is increasingly common, however, patient characteristics have higher levels of osteoporosis, co-morbidities and pain by virtue of the injury compared to OA patients (77, 78). Therefore, this thesis considers THR primarily in terms of its use for OA and idiopathic necrosis.

Hip arthroplasty entails the implantation of an artificial prosthesis to replace a damaged hip joint. The surgical procedure is broadly divided into two key components: the surgical approach, which provides access the hip joint, and the surgical technique used to replace the affected bone with the implant.

1.5.2. Surgical approach

When accessing the hip joint, several factors need to be considered including the nature of surgery (elective versus trauma), patient's age (adult or child), access requirement (femur, acetabulum and/or pelvis), the positioning of the patient, and their anatomical characteristics. Surgeon's preference and expertise, operative time and postoperative patient outcomes are also factors known to significantly influence the choice of surgical approach to the hip joint (79).

The contemporary surgical principles aim to achieve optimal access to the femoral head and acetabulum, reduce operative and bleeding time, minimise muscle and soft tissue dissection, and lower the risks of complications such as dislocation, infection and

neurovascular damage, while ensuring satisfactory postoperative pain management and early mobilisation (80). The concept of 'internerveous plane' applies to hip exposure, promoting the identification of a safe anatomical plane between adjacent nerves to reduce the risk of nerve damage. Surgical approaches to the hip can be categorised into anterior, lateral, posterior, medial, and combined, with medial approaches being rarely used due to limited access to the femoral head and acetabulum (81). Several modifications to the original approaches have been developed and are well chronicled in literature. The following paragraphs discuss the most used surgical approaches: posterior, direct lateral, direct anterior and anterolateral.

One of the prevailing techniques for total hip replacement is the posterior approach, which is also the most widely used worldwide for total hip replacement (82), and was first documented by Langenbeck and later modified by Kocher (83). The posterior approach provides excellent exposure to the acetabulum and proximal femur and spares the abductor mechanism during dissection allowing rapid postoperative rehabilitation. However, it is associated with higher incidences of perioperative sciatic nerve damages and postoperative dislocations and wound infection (79, 84-86).

The direct lateral approach, introduced by McFarland and Osborne (87) and later modified by Hardinge (88), is the second most common approach for primary THA. This approach provides extensive exposure of the proximal femur and acetabulum whist preserving vascularity and therefore has a lower risk of nerve damage. However, it is associated with a longer recovery time due to the dissection of the gluteal muscles, as well as, abductor weakness and trochanteric complications, primarily non-union and heterotopic ossification (80).

The anterior approach, initially described by Smith-Petersen during his work on mold glass arthroplasty (89), is commonly used for pelvic osteotomy, synovial biopsy and the correction of congenital dislocations of the hip; it is also increasingly used for total hip arthroplasty. A minimally invasive direct anterior approach (DAA) was developed in 1917, and showed promising results in terms of early recovery and lower morbidity (90-94). Several studies compared postoperative outcomes of primary elective THA comparing DAA to other surgical approaches and found reduced dislocation rates (85, 95, 96), better pain relief and functional recovery at 2 weeks postoperatively (97-101), with dimensioning long-term functional advantages at one-year follow up in comparison to the posterior approach (98, 99). Nonetheless, DAA remains a technically challenging procedure with limited access to the femoral medullar canal, rendering it unpopular among arthroplasty surgeons, and necessitates a substantial learning curve of approximately 100 cases for the operative time to level off (102, 103).

Watson-Jones (104) developed the anterolateral approach in 1936 with the internervous plane lying between the tensor fascia lata and the gluteus medius muscle, and this was later modified by Mueller and Charnley (105, 106). It provides adequate exposure to the femoral medullary canal with limited access of the acetabulum. The anterolateral approach provides adequate stability with only low risks of hip dislocation and sciatic nerve damage, merits similar to the DAA, but has a higher incidence of abductor muscle weakness and superior gluteal nerve damage, similar to the direct lateral approach (80). The plethora of surgical approaches used for hip arthroplasty enables a meticulous selection process that allows for the careful consideration of the unique merits of each approach, coupled with the surgeon's skills and expertise, to cater to the specific needs of the individual patient.

1.5.3. Surgical technique

Upon incising the capsule, the hip joint is dislocated providing exposure to the acetabulum and the proximal femur. Acetabular preparation involves reaming the arthritic changes and reshaping the acetabulum to accommodate the acetabular component, which is secured using cemented or press-fit (cementless) fixation with 20 degrees of anteversion. The femoral neck is carefully cut using an oscillating saw, about 1-2 centimeter above the lesser trochanter. The cancellous bone at the posterolateral corner is removed using a small osteotome, commonly known as a Box Chisel, or bone nibbler, to create an entry point with 5-10 degrees of anteversion. Two-sized T-handled rasps, known as canal reamers, are then consecutively advanced in a posterolateral fashion aimed towards the centre of the patella. Cancellous bone from the lateral femoral canal is removed from Gruen Zones 1 and 2, using a Charnley curette. Various sized modular broaches are then sequentially impacted into the medullary canal whilst applying direct pressure posteriorly and laterally. Clinically, the final broach size should provide the feeling of a 'snug-fit', aligning with the preoperative surgical plan and stem size template. The femoral canal is then irrigated followed by the placement of the femoral stem implant using either a cemented technique, with the cement restrictor inserted beforehand, or through the cementless press-fit method (107, 108). Fixation type is considered in more detail in Section 1.10.

1.5.4. Hip prosthesis zones

In view of the increased prevalence of osteolysis and loosening following THR in the 1970s, DeLee and Charnley (109) and Gruen *et al.* (110) established systems to classify the radiolucent lines at the bone-prosthesis interface seen on the anteroposterior (AP) and lateral views of radiographs. DeLee and Charnley classified the acetabular component on AP radiographs into three parts: superior, middle, and inferior, whilst Gruen *et al.* divided

the proximal femoral-prosthesis junction into seven anatomical "Gruen" zones from the greater trochanter (zone 1) to the lesser trochanter (zone 7), as shown in Figure 1-1.

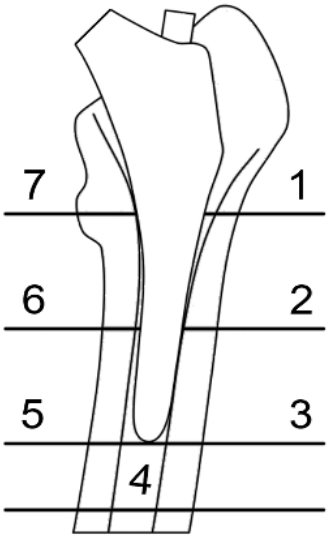


Figure 1-1 Schematic illustration showing the location of Gruen zones

Duncan and Masri in 1995 devised the Vancouver classification to standardise the reporting of femoral prosthesis complications, primarily femoral periprosthetic fractures (FPPF) and implant loosening (111). This valid and reliable system distinguishes between three types of FPPFs (A, B and C), taking into consideration fracture location relative to the prosthesis, stability of fracture, and the surrounding bone stock (Figure 1-2) (112, 113).

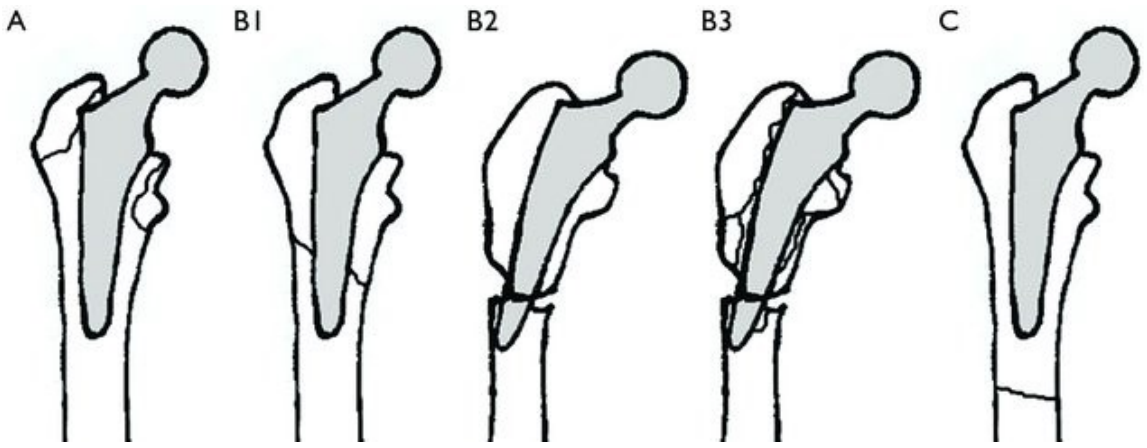


Figure 1-2 Vancouver classification for periprosthetic femoral fractures (112)

Vancouver *A* refers to fractures in the intertrochanteric area and can be subdivided into *A1* - fracture of lesser trochanter, and *Ag* - fracture of greater trochanter. *B* cases are diaphyseal fractures around or just below the femoral stem: *B1* – stable stem, *B2* – loose stem, and *B3* – loose stem associated with significant bone loss. Vancouver *C* are diaphyseal fractures below the femoral prosthesis.

The literature and prevalence of femoral periprosthetic fracture are described in Chapters 6 and 7 whilst periprosthetic fracture risk during femoral press-fit broaching and implantation is investigated in Part II of this thesis.

1.6. Hip Resurfacing Arthroplasty

1.6.1. Initial development

Hip resurfacing arthroplasty (HRA) remained unpopular until the early 1990s. The advancements in understanding bearing surfaces and metal-on-metal (MoM) articulations allowed trials of several implant design modifications. Wagner's press-fit carbon cobalt chrome Molybdenum design, trialled primarily on female patients with a history of dysplasia, yielded a significant early revision rate (14.3%) due to loosening or fracture (114). McMinn's cementless MoM cobalt chrome femoral component and acetabular shell showed 8.6% revision rate at 54 months and around 63% survival at 16 years (74). Further modifications to McMinn's design through the addition of hydroxyapatite (HA) at bearing interfaces or cementation of both femoral and acetabular components led to similarly poor implant failure outcomes (115). The final evolution of McMinn's design, the Birmingham Hip Resurfacing method by Smith & Nephew in the 1990s, included a hybrid cemented femoral stem and press-fit acetabular cup coated with HA and improved 10-year survival to 92-96% (116-120).

Harlan Amstutz, who is known for developing eccentric internal shells and developed press-fit designs between the 1960s and 1980s, enhanced his HRA Conserve Plus implant design in the USA and reported 88.5% survival at 10 years (121). The resurrection of HRA in the early twenty-first century led a few manufacturers to introduce their own implants, two examples are ASR by DePuy and Durom by Zimmer. Notable variations in implant design and materials used contributed to contrasting outcomes and subsequently the withdrawal of the implants due to the high failure rate, as reported in registry data (116, 122). The concept of hip resurfacing, then, received negative publicity and widespread concerns emerged regarding early failures associated with the use of metal-on-metal (MoM) bearing surfaces in HRA (123). Several reports documented adverse reactions to metallic wear debris and increased incidence of soft tissue lesions, pseudotumour, in the periprosthetic resurfaced hips and raised metal ion levels, also known as metallosis (124-128). However, evidence shows excellent outcomes in carefully selected patients receiving BHR and Conserve Plus prostheses with ten-year survival rates of 96.3% and 98.7%, respectively (129, 130).

1.6.2. Patient selection and indication

Whilst total hip replacement offers excellent outcomes for patients with end-stage hip osteoarthritis, its design arguably means that patients have a limited potential to return to high-level physical activity, with additional risks of wear and revision surgery for young patients (131, 132). HRA aims to mimic native anatomy and preserve bone stock, thus, recreating patient's normal gait and range of motion (133-135). However, patient selection remains a topic of debate in the orthopaedics community with limited consensus on the ideal patient to benefit from HRA. Several independent patient-reported factors were reported to be associated with early failure in HRA including female gender, small femoral head (of less than 50 mm), low bone mineral density and dysplastic hips (124).

The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) reports higher revision rates for females under 55 years of age compared to males (18.6% versus 4.6%) (7), figures echoed by the International Hip Resurfacing group at 10 years (90% versus 99%) and 21 years (81.3% versus 92.5%) (136). Additionally, female gender is thought to be compounded by the other three negative prognostic factors mentioned above (124, 132, 137). Increasing femoral head size (>50 mm) was shown to be associated with lower revision risk (cumulative revision risk at ten years: 44 mm size ~ 20% versus 50 mm ~ 10%) (6), which is likely due to the stress shielding effect related to the femoral peg static dimension of older HRA implant designs when implanted in diminutive femoral necks (121, 122, 138).

Although there is no established bone mineral density (BMD) threshold for hip resurfacing, biomechanical testing on cadaveric specimens showed that BMD below 0.65 g/cm^2 exhibits a higher risk of neck of femur fracture following HRA (139, 140). Similarly, a series of 447 hips showed lower revision risk for HRA due to OA in comparison to rheumatoid arthritis (RA) or avascular necrosis (AVN) (116, 130). Lastly, patients with dysplastic hips, also known as developmental dysplasia of the hips (DDH), were shown in the AOANJRR annual report to have an increased failure rate compared to primary hip OA (16.3% versus 7.2%) (7), and is partly due to the shallow dysplastic morphology of the femur and more inclined acetabulum (116).

Therefore, current evidence recommends the following patient-related factors for MoM HRA:

- physically active male patients;
- below 55 years of age;
- diagnosis of primary hip osteoarthritis;

- adequately sized femoral head (>55 mm); and
- no prior history of DDH or conditions of low BMD (such as AVN or RA).

In addition to the abovementioned implant- and patient-related factors, surgeons' expertise is a key factor for HRA success and is associated with a steep learning curve averaging a threshold of fifty cases for non-specialist hip surgeons, equating training for six months in a high-volume centre, to minimise the risk of complications and enhance patient reported outcomes (PROMs) (141). For experienced hip surgeons, the threshold to avoid early complications was found to be twenty-five procedures (142). The use of computer-assisted navigation systems is thought to further reduce the learning curve (143), however, a recent study showed no additional value in terms of femoral implant positioning and PROMs at one-year postoperatively (144).

1.6.3. Subsequent development of HRA

Given the ongoing concerns related to the MoM bearing surfaces and the associated soft tissue metallic debris reactions, manufacturers' interest to develop alternative bearings has surged in recent years. In theory, the use of ceramic-on-ceramic (CoC) or metal-on-polyethylene (MoP) may prevent the metallic reactions of MoM HRA and may extend its application to female patients and those with smaller femoral head sizes (<50 mm).

Previous attempts at devising CoC HRA failed due to metallic debris, liner fracture or accelerated wear of the ceramic femoral head coating (145, 146). At present, two novel cementless CoC implant designs for HRA, H1 (Embody Orthopaedic; Imperial College London, UK) and ReCerf (MatOrtho, UK), are being evaluated in clinical trials. The current H1 prosthesis yields potential in mitigating the aforementioned risks through its monobloc design, coated with HA and plasma-sprayed titanium, along with utilising advanced ceramic material, BIOLOX delta.

Similarly, the development of MoP bearing resurfacing has been incremental, with earlier implants that used conventional polyethylene failing due to volumetric wear, neck of femur fracture, osteolysis or femoral component aseptic loosening (132, 147, 148). McMinn's Polymotion hybrid implant (JointMedica, UK) comprises of a cemented metal femoral head and a cementless highly cross-linked polyethylene (HXLPE) coated with plasma-sprayed titanium acetabular component. Despite being in its early stages, preliminary clinical findings of eighty-four patients are promising and reported only a single incidence of head-neck junction lucency with no other documented complications of revision, fracture, osteolysis or component migration (149).

1.7. Study of outcome in joint replacement

Implementation of evidence-based practice in arthroplasty is challenging and requires a robust understanding of the different study designs, their susceptibility to bias, and the interplay of various patient-, surgical- and implant-factors to guide clinical decisions. The ability of a study to evaluate outcomes depends on the chosen design and the quality of methods. The hierarchy of study designs based on their inherent susceptibility to bias and confounding was initially described by the Canadian Task Force in 1979 (150). The Centre for Evidence-Based Medicine in Oxford established the widely used levels of evidence (151), which was later adapted by the orthopaedics community (Table 1-1) (152-154).

Levels of evidence		Type of studies
Controlled trials	Level I	High-quality randomised controlled study (RCT) Systematic review or meta-analysis of level 1 RCTs
Observational (Analytical or descriptive) Except for poor-quality RCTs	Level II	Prospective cohort study (including computerised medical records) Poor-quality RCTs Systematic review of level II studies
	Level III	Case-control study Retrospective cohort study Systematic review of level III studies
	Level IV	Case series
Descriptive or basic science	Level V	Case report Expert opinion Descriptive studies with no control group Animal or cadaver studies

Table 1-1 Levels of evidence in orthopaedics (152, 154)

Nevertheless, the choice of study design is influenced by external factors such as feasibility of intervention, ethical considerations, time and financial cost. Advances in recent years in data analytics, data mining and machine learning have fuelled new types of evidence, broadening the contemporary levels-of-evidence hierarchy (155, 156). Furthermore, examples of clinical adaptive retrospective trials in cancer research receiving regulatory approvals have set precedents for achieving accurate conclusions quicker whilst avoiding the lengthy process needed for standard prospective RCTs (157, 158).

Several study types are employed in this thesis, primarily levels II, III and V, and are delved into throughout Chapters 3 to 9. This section provides a brief overview of the different study designs used in orthopaedic studies whilst the outcome metrics used in joint replacement are addressed in Section 1.8.

1.7.1. Controlled studies

Randomised Controlled Trials (RCT), if well-conducted, remains the study of choice for evidence-based practice. The processes of randomisation and blinding in RCTs address issues related to systematic bias and eliminate the influence of confounding variables (159). However, several barriers to implementing a rigorous RCT exist in surgical practice, including ethical constraints; difficulties in randomisation and blinding to surgical interventions partly attributed to patient preference and the lack of clinical equipoise; time and financial cost; and external validity concerns related to heterogenous disease effect (160).

Recent evidence indicates marked improvement in the design and methodological rigour of clinical trials in orthopaedic surgery (161, 162). Though, these trials still face significant challenges concerning complex recruitment pathways and logistical issues, with the majority being conducted in expert centres (163). Similarly, the choice of outcome metrics in joint replacements may present a substantial barrier to the implementation of clinical trials (see Section 1.8).

Safety outcomes, such as mortality and implant failure, are common measures in arthroplasty requiring a large number of patients and lengthy follow-ups to demonstrate significant variations in implant selection. The growing emphasis on patient-centred indicators is expected to further increase the adoption of RCTs. However, this shift also necessitates careful attention to limiting bias related to the variability of subjective measures and determining the minimal clinically important difference (MCID) for such measures (164, 165).

1.7.2. Observational studies

Whilst RCTs remain the gold-standard, observational studies form the vast majority of the orthopaedics literature. Coupled with advancements in data analytics, they are influential in joint arthroplasty guidelines (166). Observational studies encompass two main categories: *analytical* (level II design: prospective cohort, level III designs: retrospective case-control or cross-sectional), and *descriptive* (level IV design: case series).

The lack of randomisation and blinding in observational studies render them susceptible to systemic bias and confounding factors. Systemic bias in observational studies often relate to patient selection and data collection methods, reducing the generalisability of findings. Confounders, on the other hand, hinder the internal validity and may lead to, if not accounted for, distorted association between the exposure and dependent variables. Strategies to minimise confounding effects include *restriction* by employing well-defined inclusion criteria, *matching* techniques to control for confounding by indication, and *statistical control* methods to adjust for the influence of confounders (167). Despite the use of these methods, there may remain residual or unmeasured confounding (168).

Case series form the second pillar of observational studies, and the descriptive collection of case reports involving patients with similar conditions or treatments are commonly used to report rare adverse events. In arthroplasty, they commonly provide a source of information on the safety and durability of newly approved implants. However, case series lack the scientific rigour, are subject to observer bias and lack hypotheses and the ability of inferential testing.

1.7.3. Joint Registries

The mission of national joint registries (NJR) is to monitor implant performance and provide timely warnings through capturing longitudinal data on implants' safety,

effectiveness, and failure rates (169, 170). Joint registry studies are a subset of observational studies (levels II or III) specifically focused on analysing the outcomes of large numbers of patients who undergo joint-related procedures, primarily arthroplasty.

The Swedish Hip Arthroplasty Registry (SHAR) was founded in 1979 and is the world's first national hip register collecting data on primary hip arthroplasty. Three Nordic countries subsequently followed the Swedish NJR movement; Finland (1980), Norway (1987) and Denmark (1995) whilst Australia and the New Zealand both started collecting data in 1999 (154). The establishment of the National Joint Registry for England and Wales (NJR) took place in 2003, and is currently the largest orthopaedic registry in the world (6).

Advantages of NJR data include its large sample size of longitudinal observations documented over extended periods, which in most cases can be linked to the mortality register. NJR provides a valuable source for comparative effectiveness studies and health economics research at a population-level across diverse demographic groups and regions. Similarly, the large database incorporates diversity of patient selection, surgical techniques, and implants, and thus allows the study of rare events and the ability to conduct subgroups analyses.

However, it is important to note the limitations of NJR investigations and what they cannot achieve. First and foremost, NJR studies are observational and hence may not satisfy the Bradford Hill criteria in establishing causal inference (171) and are also subject to the biases of observational studies discussed in Section 1.7.2. Most importantly, registry databases were established for the purposes of implant safety surveillance, rather than patient safety, using revision rate as a primary endpoint; the limitations of using this outcome measure are discussed in Section 1.8. Despite improvements in data collection, NJR data lacks detailed information on health co-morbidities and has substantive

incomplete (or missing) data. Selection bias is another limitation, wherein only patients eligible or receiving joint replacements are reported, which restricts the generalisability of findings and may exacerbate disparity (172).

Part I of this thesis predicts and evaluates outcomes from the Swedish Hip Arthroplasty Register (SHAR) with external validation analyses using the National Joint Registry (NJR) for England and Wales; both are described in Chapter 2.

1.7.4. Laboratory studies

Unlike controlled trials and observational studies, laboratory studies serve as a foundation for the development of new interventions but require clinical validation to translate findings to patient care. They are particularly helpful to enhance the understanding of certain outcomes whilst providing a path for bidirectional translation. Animal or cadaveric studies often rely on controlled testing environments to manipulate factors that affect the outcomes of interest. This provides a platform for hypothesis generation and testing. However, the major drawback is their limited ability to directly translate findings to real-life settings. Likewise, the tested specimens might not fully mimic living human biological or physiological processes, thus minimising the generalisability of results. Furthermore, laboratory studies are labour-intensive and cost-consuming whilst demanding specialised expertise in the relevant disciplines to address the multifaceted research questions.

Part II of this thesis consists of two biomechanical laboratory studies investigating bone strain behaviour during femoral cementless fixation to generate an understanding of periprosthetic femoral fracture pattern using synthetic and cadaveric study designs.

1.8. Outcome metrics in joint replacement

National Joint Registries (NJRs) primarily focus on collecting measures related to the monitoring of implant performance but lack data pertaining to the wider spectrum of patient safety. Four levels of data can be collected by NJRs based on the type of information collected (Figure 1-1), leading to classifying joint registries into a hierarchical tier system (173, 174).

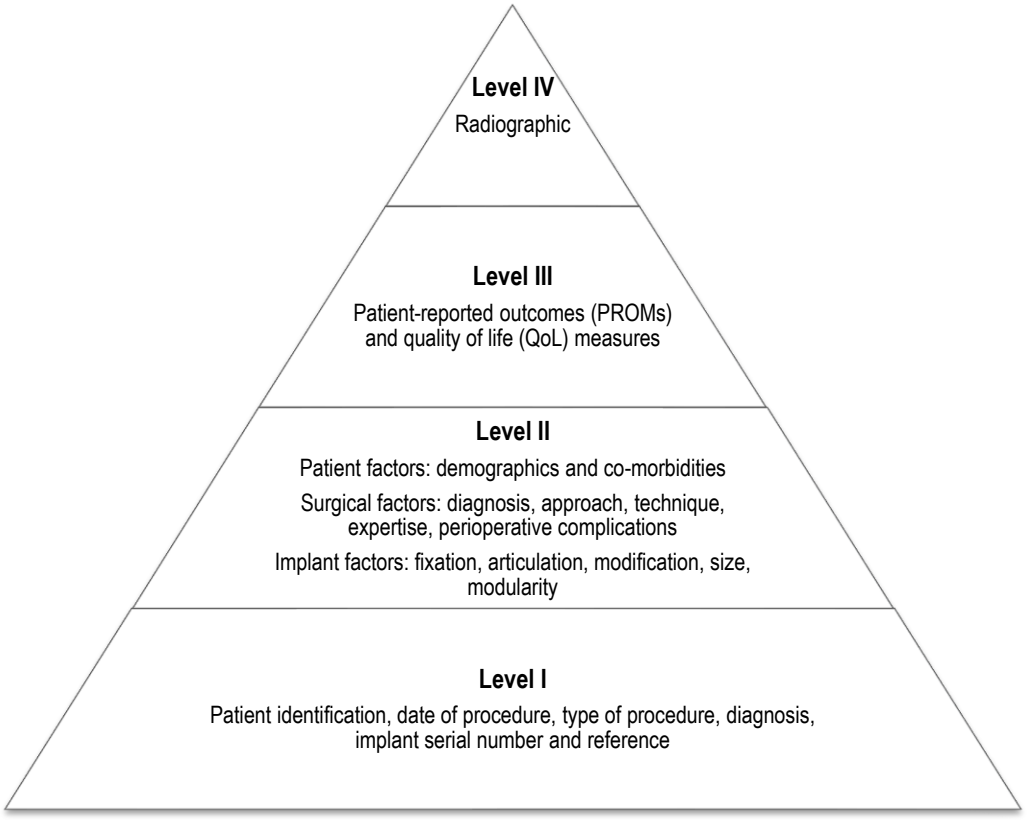


Figure 1-3 Levels of data collected by NJRs

Level I data includes patient identifiers, diagnoses, and basic procedural data, which on its own may permit the study of implant-specific revision rates. Level II data includes additional information related to patient factors; demographics and co-morbidities, surgical factors; surgical approach and technique, surgeon expertise and perioperative complications, and implant factors; fixation type, articulation, modification, size and modularity. Most NJRs rely on these two levels to evaluate associations including implant

performance, perioperative complications and hard endpoints patient outcomes (implant failure or mortality).

Patient-reported outcomes (PROMs) are considered level III data and can be used to assess subjective outcomes following joint replacement such as symptoms, functional status and satisfaction. PROMs data relies on patient input and hence are costly to collect and is associated with a higher percentages of missing data points. Level IV data includes radiographic information and can be used to study subclinical implant failure, implant positioning, migration, alignment, and long-term wear and osteolysis. Collection of level IV data requires extensive computational power and storage, compliance by units and centres, as well as significant financial expenses, thus they are used less often (173). To achieve a comprehensive understanding of arthroplasty performance, it is key to evaluate a combination of relevant outcome measures.

1.8.1. Mortality

Surgical procedures, including joint replacements, carry a risk of postoperative death, which is used as an endpoint measure. Postoperative death is defined as the occurrence of patient death within a specified timeframe following a surgical procedure. In primary elective hip arthroplasty, 90-day mortality risk is reportedly low ranging between 0.4% (NJR) and 1.9% (SHAR) (5, 6). The rare occurrence of mortality, whilst paramount to aid the clinical decision-making process, is difficult to detect a significant difference and requires a large sample size particularly in the context of RCTs.

Kaplan-Meier (KM) estimates, and Cox proportional hazard regression are commonly used to evaluate mortality outcomes. Kaplan-Meier estimates are used to represent the cumulative survival curve whilst Cox regression analysis allow comparisons among patients, procedures, and implants. However, KM's non-informative censoring

assumption implies that the probability of no-event is independent of survival time and can lead to biased survival estimates if not adjusted for.

Similarly, Cox regression analyses assume linearity and constant hazard and that can be misleading in cases of association between predictor factors and hazard rate over time. Consider a hypothetical study comparing the survival rates of two types of hip arthroplasty: total hip replacement and hip resurfacing arthroplasty. The Cox regression assumptions mean that the hazard of mortality changes uniformly over time, masking differences in hazard rates between the two interventions. Whilst in reality, each intervention might have distinct hazard patterns. Newer statistical methods such as penalty regularisation for non-essential variables may help differentiate the pertinent factors affecting each intervention (175).

Finally, arthroplasty-related mortality may not be directly attributed to the intervention itself but rather might be confounded by pre-existing medical conditions or surgeon and anaesthetist expertise. Despite the importance of this crude measure as a safety outcome, caution is required when interpreting the findings.

1.8.2. Implant survival

Whilst mortality measures holistically assess the safety of surgical interventions, implant survival evaluates implant performance through the occurrence of re-do surgery. Historically, implant survival has served as a principal endpoint metric in joint registries and encompasses two fundamental metrics: re-operation and revision. Re-operation is defined as any additional surgery to the joint, regardless of the actions taken to any of the implant components (replaced, extracted or left untouched) whereas revision is defined as a re-operation where at least one component (for example in the hip; femoral, liner or acetabular components) is exchanged, extracted from or added to the prosthesis (169).

In primary hip replacement, cumulative revision risk ranges from 3% (NJR) to 8.8% (SHAR) (169, 170). Similar to mortality metrics, revision outcomes in arthroplasty are subject to deficiencies related to their relatively rare occurrence and the use of classical statistical models discussed in the previous sub-section.

The use of revision rates as a sole endpoint has additional limitations. First of all, the decision to reoperate is multifaceted and includes factors related to patient co-morbidities, symptoms, satisfaction and functional outcomes; radiographic assessment; choice of revision type; and surgeon and unit expertise, all of which are not captured in registry datasets. Likewise, when compared to 1-year revision risk (8.24% in NJR) (6), previous studies estimated that up to 20% of patients are not satisfied after their hip arthroplasty (176). Taken together, survival analysis of the revision rates as a mean to establish implant performance can lead to biased results.

Finally, censoring observations when the occurrence of death preceded a revision surgery is interpreted as successful cases in terms of implant survival and may alter (or bias) the precision of estimates. Whilst attempts to minimise this effect through competing-risk analyses ("1 - Kaplan-Meier estimate"; using death as a competing risk for revision hip arthroplasty), its use can be misleading and attenuate the probability of revision (177).

1.8.3. Measures of functional outcome

Level III data includes functional outcomes, which can be *generic* concentrating on quality of life (QoL) and general well-being; *condition-specific* focusing on symptoms of a particular disease; or more reliably a combination of the two (178). Likewise, reporting of functional outcomes can be either patient-reported, observer-reported or both. However, this method of data collection is costly and relies on patients' and assessors' compliance and hence is underreported in national registries (179, 180).

General quality of life (QoL) instruments includes the European QoL 5-Dimensions (EQ-5D), Short Form 12 (SF-12), Short Form 36 (SF-36) and the Visual Analogue Scale (VAS). These tools use standardised questions to evaluate the individual's physical and mental wellbeing, pain, and functional levels. EQ-5D is composed of five domains in addition to a separately recorded VAS (EQ-VAS) for general health. The domains are: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, with each domain divided into three severity levels.

SF-36, and its shorter version SF-12, assess additional dimensions related to social and emotional functioning and each generate into physical (PCS) and mental component summary (MCS) scores. Whilst SF-12/36 scores range between 0 (worst) to 100 (best), EQ-5D scores can be reported in two ways: health profile of domain scores or global index weighted to health states ranging between -0.594 (state worse than death) to 1 (best possible health state) (181, 182). VAS scales commonly range between 0 and 100 and can be used as a generic score, as in EQ-VAS, or specifically to measure indicators such as pain intensity, satisfaction and anxiety levels (181). All the above mentioned four scores have demonstrated satisfactory validity, reliability, and sensitivity (183-187) and significant agreement between instruments (188-190).

On the other hand, patient-reported outcome measures (PROMs) are specific and ask questions to evaluate symptoms and function. For the hip, besides VAS-specific scales, the most commonly used instruments include: the Harris Hip Score, the Charnley classification, the Western Ontario and McMaster Universities Arthritis Index, and the Oxford Hip Score. The Harris Hip Score (HHS) was developed in 1969 as a hip-specific tool aimed to assess outcomes following hip procedures. HHS is a physician reported tool consisting of ten questions across four domains: pain, function, absence of deformity and range of motion (191). Despite its widespread use, its ceiling effects may lead to deficient

discriminative ability to detect the minimal important difference (MID) in patients with mild or severe hip conditions (192). On the other hand, the Charnley functional PROM system, initially described by Sir John Charnley in the 1970s, assigns patients to one of three categories using two patient-administered questions; A - unilateral hip disease; B - bilateral hip disease; and C - multiple joint disease or major medical conditions hampering mobility. Despite the reported inconsistencies in using the Charnley classification, it was shown to be a robust predictor of PROMs and permits adjustments to varying musculoskeletal co-morbidities (181, 193).

The Western Ontario and McMaster Universities Arthritis Index (WOMAC) was developed in 1982 and is a valid and reliable osteoarthritis-specific index consisting of twenty-four questions measuring symptoms related to hip and knee osteoarthritis across three domains: pain, stiffness, and physical function (194-196). The WOMAC is widely used to evaluate PROMs following total hip arthroplasty (197) but when compared to the Oxford Hip Score (OHS, see below), the WOMAC had a lower effect size to detect changes following THR and exhibited a diminished response rate partly attributed to its extensive nature (198).

The OHS was introduced in the 1996 predominantly to evaluate PROMs changes in clinical trials (199), and is the most commonly used hip procedure-specific instrument in the United Kingdom (200). OHS uses a valid and reliable twelve-item questionnaire related to pain and functional ability, each allocated a score between 0-4 (original scoring 0-5), with a summary score ranging between 0 (worst) to 48 (best) (201, 202). However, Beard *et al.* identified the minimal important change (MIC) to be approximately 5 points (Δ >10% from baseline) (203), with a recent study reporting MDC for observational cohort studies to be 12.4 points (Δ >25% from baseline) highlighting its limited ability at detecting

subtle changes over time (204). Table 1-2 summarises functional outcome measures used in hip replacement.

Outcome measure	Type	Reported by	Range
EQ-5D	General QoL	Patient	-0.594 - 1
SF-36	General QoL	Patient	0 - 100
SF-12	General QoL	Patient	0 - 100
VAS	General or specific	Patient	0 - 100
Charnley Classification	Hip Specific	Patient	A,B,C
HHS	Hip Specific	Assessor	0 - 100
WOMAC	OA Specific	Patient	0 - 96
OHS	OA and Hip Specific	Patient	0 - 48

Table 1-2 Key functional outcome instruments used in hip replacement

A qualitative study evaluated patients' perspective regarding the OHS (205). Whilst the study had a limited degree of theoretical saturation and restricted capacity to generalise the findings, the patients indicated that the OHS fell short in addressing crucial aspects related to living with a hip replacement, primarily due to its narrow emphasis on functional and physical health:

"I have a daughter who lives on the 4th floor, and I keep myself from visiting her because the stairs are a nuisance for me. Alternatively, I must take pain medication before I go, and then it sort of works when I get there, and then I can climb the stairs"

(Male, 67 years, preoperatively)

"The fact is that I really want to go for a walk but when I am done walking… then I am completely worn out when I get home"

(Female, 71 years, 12 months postoperatively)

Although PROMs provide added advantages over hard endpoint outcomes (such as mortality and implant survival) in evaluating outcomes following orthopaedics interventions, they are subject to confounding factors and ceiling effects. Fluctuations in PROMs may be influenced by factors related to patients' age, overall health and other joint pathologies (206, 207). A recent study investigated the use of the metabolic equivalent of task (MET) score, a personalised numerical measure of physical activity, to overcome the ceiling effect associated with conventional PROMs (208). Whilst further research is required, the authors demonstrated additional gains beyond the maximum OHS score in relatively young patients post-primary THR and HRA. Nevertheless, uncertainty remains when attributing changes in PROMs to a particular intervention.

1.8.4. Objective measures

Objective measures refer to the consistent, valid, and reproducible data assessing a specific phenomenon or outcome. For joint health, they represent level IV data and can be primarily divided into radiological and biomechanical measures. Radiological assessment following hip arthroplasty can be quantitative or qualitative, evaluating factors related to prosthesis positioning, alignment, wear, and complications such as periprosthetic fracture or loosening.

Whilst radiological measures allow for longitudinal tracking of structural changes and treatment outcomes, they can be associated with significant costs and possible radiation exposure risks, and they failed to capture information related to patient-reported and functional outcomes. On the other hand, gait analysis measurements using wearable sensors, motion capturing systems or treadmills, are able to quantify differences in gait patterns pre- and post- hip arthroplasty (209-213). In contrast to the three other endpoint outcomes (mortality, revision, and PROMs), radiological and gait data can serve as

surrogate markers, thus, a comprehensive understanding of joint health requires a harmonic integration of multiple outcome measures.

1.9. Factors affecting outcome in primary hip arthroplasty

A multifaceted interplay of various determinants influences the extent of outcomes following arthroplasty and can be divided into three overarching domains: patient factors, surgical factors, and implant-specific factors (Figure 1-2). An extensive body of research exists in evaluating these domains for primary hip arthroplasty (12, 96, 214-217). A UK NJR-study compared an exemplar hospital that uses a single type of total hip replacement and is associated with higher survival rates with other hospitals in the country; and showed "better than expected" implant performance is associated with implant selection only (218). This alludes to the importance of modifiable implant factors to evaluate survival outcomes.

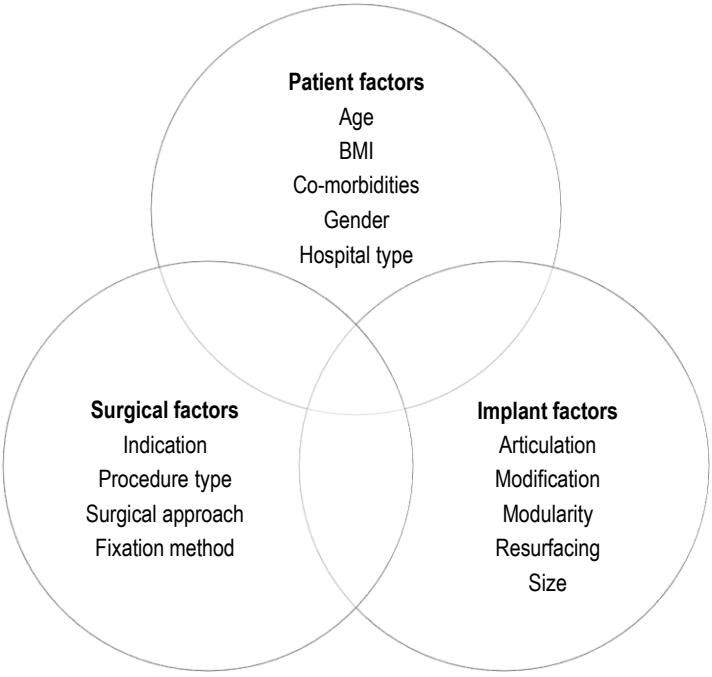


Figure 1-4 Factors affecting arthroplasty outcomes (216)

1.9.1. Patient factors

The effects of patient characteristics on outcomes following total hip replacement are well-chronicled in literature. Despite this, the effect size of each remains controversial.

1.9.1.1. Age

Increasing age is an established risk factor for premature mortality following THR (12, 219-223). The causes of perioperative death, despite adjustment for co-morbidities, are varied and most commonly include ischemic heart, cerebrovascular, and pulmonary disease, diabetes, cancer and multiorgan failure (223). Analyses of registry data from Australia, Sweden and the United Kingdom reports inferior survival, and lower revision risks, in elderly patients after primary hip replacement (5-7, 224, 225). Population-level research suggests underlying aging physiological processes contribute to the increase in mortality risk irrespective of disease history (226). In contrast, there is strong evidence to suggest that youth predisposes to revision surgery, or multiple re-revision procedures, following THR (227). This is partly due to the likelihood of exceeding the lifetime risk of revision in younger patients (227, 228).

Evidence investigating the effect of age on functional outcomes is contrasting with some studies reporting equal hip-specific and general QoL outcomes following THR (229-232) whilst others favour greater benefits for the younger populations (181, 233-236). However, variations in implant choice, functional outcome tools, timepoint of assessment, co-morbidities (if not adjusted for) and dichotomisation of age into groups, are all factors that might affect differences in functional outcome scores. Likewise, unrealistic young patients' expectations may also diminish their postoperative PROMs after THR (237-239). A recent study categorised age into three different groups and evaluated QoL (EQ-5D and

SF-36) and hip-specific (WOMAC, HHS and HOOS*) measures whilst adjusting for baseline activity levels up to five years following THR using the same implant (240). The authors reported no difference in postoperative functional outcomes when accounting for confounding factors. Nevertheless, conventional statistical models used in these studies might overlook the intricate relationship among the varied patient-, surgical- and implant-predictor factors.

1.9.1.2. Body Mass Index

The relationship between body mass index (BMI) and mortality after THR has yielded inconclusive findings in the literature. Overweight, which is defined as a BMI ≥ 25, was found to have a protective effect trend on mortality outcomes, a term previously coined as the obesity paradox (241, 242). Hunt *et al.'s* (12) analysis of modifiable perioperative factors associated with 90-day mortality post primary hip replacement in the National Joint Registry for England and Wales identified obesity as a protective factor in terms of mortality (HR 0.76, 95% CI 0.62–0.92, p = 0.006). This trend was in agreement with previous studies (243, 244), however, other authors found no significant association with mortality (219, 245, 246). Huddleston *et al.* reported an increased risk of adverse events in obese patients (OR 1.20, 95% CI 1.02–1.42, p < 0.032) but no statistically significant association with mortality (247). Similarly, other retrospective studies reported varying levels of positive associations between higher BMI classes and mortality risk; the *Ontario administrative health care database* (n = 22,251, BMI > 45 kg/m^2; RR 1.38 95% CI 1.18–1.62, p = 0.07) (248); the *German healthcare insurance database* (n = 131,576, BMI > 40 kg/m^2; OR 1.8; 95% CI 1.3–2.6, p < 0.001) but no relationship for BMI levels between 30 and 39 kg/m^2 (249); whilst a study using the *South Korean National Health Insurance Service dataset* (n = 3,627) reported that both overweight (BMI 25–29.9 kg/m2; HR 2.0; 95% CI 1.0–4.0, p value

* Hip Disability and Osteoarthritis Outcome Score

not reported) and underweight (BMI under 18.5 kg/m2; HR 2.3; 95% CI 1.1–4.9, p value not reported) are associated with an increased mortality risk in THR (250). The latter, being underweight, was demonstrated to have an eight times increased risk of two-year mortality following total hip and knee replacements (251). Other authors found no relationship between BMI categories and mortality risk after THR (n = 809, HR 1.06; 95% CI 0.96–1.17, p = 0.25) (252). A recent meta-analysis of fifty-three heterogenous studies showed no significant association between BMI and mortality from 30 days to 10 years after primary elective hip arthroplasty (223). Obesity, however, is known to be associated with longer operative time, bleeding risk, and the development of co-morbidities such as coronary heart disease and hypertension which impact outcomes as discussed in the following sub-sections (253-255).

With regards to implant failure outcomes, unlike mortality, there is sufficient evidence to suggest increased risk of adverse events in patients with high BMI (256). Obese patients, those with BMI ≥ 30 kg/m2, were shown to have statistically significant higher independent risks of infections, dislocations, re-operations, revisions and re-admissions (255-260). On the other hand, there is a paucity of controversial evidence evaluating postoperative outcomes for underweight patients undergoing THR, which can partly be explained by the relatively low proportion of said patients. A recent study using a large insurance claims database of over 50,000 patients in the United States matched underweight patients to a control cohort of normal BMI and reported higher incidences of revision, periprosthetic fractures and sepsis (261). This is in contrast to previous smaller-scale studies describing a significant increase in length of stay and bleeding risk for underweight patients but no statistically significant differences related to postoperative complications (262, 263).

Functional outcomes for different BMI classes in the literature appear to favour no significant long-term differences. A few studies reported lower pre- and post-operative functional outcomes (using OHS, HHS, SF-36 and/or WOMAC) for obese versus non-obese patients, however, no significant difference was noted in terms of the absolute gain in functional scores (264-266). A meta-analysis of six prospective cohort studies on the effect of obesity on outcomes following THR reported no difference in OHS ($n = 2$, $p = 0.60$) and minimal non-significant difference for HHS ($n = 6$, $p = 0.07$) between the obese and non-obese groups (255). Similar findings were recorded for short- to mid-term studies using OHS and SF-12 (264, 267). Long-term studies with a minimum of ten-year follow up showed no difference in functional outcomes, when evaluating a variety of OHS and HHS, between high BMI and normal BMI patients (268-270). Most recently, a longitudinal nine-year observational study noted no statistically significant differences in HHS and pain scores at one- and five- year follow up post THR among any weight group (271).

1.9.1.3. Co-morbidities

Co-morbidities are generally represented by validated indices, such as the American Society of Anaesthesiologists (ASA) risk score, Charlson Co-morbidity Index (CCI), and Elixhauser Risk Score (ERS).

There is strong evidence to suggest increased risk of mortality for patients undergoing primary THR with worse preoperative co-morbid status, particularly ASA/CCI > 3 and a higher Elixhauser score (219, 241, 244, 272-274). A history of cardiovascular disease or moderate to severe liver disease are established risk factors and are reported to increase the mortality risk by eight- and ten-fold, respectively (12, 275, 276). Other medical conditions statistically significantly associated with increased 90-day mortality risk following THR are: metastatic cancer (HR 3.14), congestive heart failure (HR 2.11),

dementia (HR 2.04), renal disease (HR 1.98), psychosis (HR 1.85), cerebrovascular disease (HR 1.40) and chronic pulmonary disease (HR 1.32) (245).

In contrast, the evidence regarding co-morbidity indices and revision risk is moderate. A higher co-morbidity index was observed to be associated with a higher revision risk at one- and five-year post THR, primarily due to peripheral vascular disorders, metastatic cancer and psychosis, however, the limited number of patients with these conditions reduced the generalisability of findings (277). This is in agreement with a meta-analysis reporting higher incidences of *revision* in patients with high blood pressure (OR 1.11; 95% CI 1.02 – 1.21) and liver disease (OR 1.96; 95% CI 1.16 – 3.30); and *surgical site infections* mainly in patients with diabetes or liver disease (278). Nonetheless, there remains a need for prospective clinical trials with a large sample size to adequately assess the severity of co-morbid conditions on implant performance outcomes.

The effect of co-morbidity indices on pain and functional outcome changes appears to be less important. Whilst higher levels of co-morbidities have worse levels of preoperative and postoperative pain and functional scores related to primary elective THR, the relative improvement from baseline status is similar across groups ($p_{pain} = 0.07$, $p_{OHS} = 0.10$) (279). Other studies reported similar findings indicating that all patients, irrespective of their co-morbidities status, have comparable improvements in pain scores, QoL and hip-specific measures following THR (280, 281). Condition-specific analyses show worse postoperative joint-specific (WOMAC) and QoL (SF-12) scores following total hip or knee replacements in patients suffering from heart disease (OR_{WOMAC} 1.24, OR_{SF-12} 1.49), lung disease (OR_{SF-12} 1.26), depression (OR_{WOMAC} 1.69), or stroke (OR_{WOMAC} 1.32) whilst postoperative pain scores were not statistically significant in relation to co-morbidity (278).

1.9.1.4. Gender

Historically, male gender appeared to predispose to premature mortality following primary elective THR (241). The NJR for England, Wales and Northern Ireland reports higher short- and long-term mortality rates for men, whilst the AOANJRR and SHAR do not publish mortality-specific gender comparisons. However, recent studies suggest otherwise. A meta-analysis of fifty-three international studies of varying levels of heterogeneity and risks of bias, showed no significant association between gender and mortality risk following primary elective THR ranging from 30 days to 10 years (223). In contrast, a retrospective study of 418,885 patients using the American College of Surgeons National Surgical Quality Improvement database reported lower 30-day mortality risk for men (OR 0.70; 95% CI 0.51-0.98; $p < 0.03$), but significantly increased risks of other adverse events including re-admission, re-operation, wound infection and length of stay (282).

Despite this, there is established evidence discerning no difference between genders in relation to implant survival for primary hip arthroplasty after adjusting for confounding factors (227, 283-285). SHAR and AOANJRR report comparable cumulative percent revision figures of primary hip arthroplasty, whilst the UK-NJR reports slightly higher, but insignificant (227), risks for females from five years postoperatively onwards (5-7).

Female gender appears to be a predictor factor for worse postoperative pain in THA (286) and lower preoperative functional scores (OHS and WOMAC) but equivalent, or greater, functional improvement compared to males at one-year postoperatively, after adjusting for baseline characteristics (181, 284, 287). Data from the Dutch Arthroplasty Register shows similar conclusions with significantly higher postoperative functional improvement in females as measured by OHS, EQ-5D, HOOS, and pain relief (288).

1.9.2. Surgical factors

Several surgical factors have been shown to influence postoperative outcomes of total hip arthroplasty including: surgical indication, preoperative planning, surgeon and unit volume and expertise, and the choice of surgical approach (241, 289-293). However, some of these factors are difficult to measure in NJRs and a recent study highlighted the importance of implant selection, rather than surgeon or unit factors, in achieving better than expected outcomes (218). The sub-sections below address the surgical factors relevant to this thesis whilst implant factors are discussed in Section 1.9.3.

1.9.2.1. Surgical indication

Whilst total hip replacement (THR) is primarily used to treat end-stage osteoarthritis, it is also used as treatment modality for other conditions such as inflammatory arthritis, trauma, metastatic bone disease (MBD), and developmental dysplasia of the hip (DDH) (see Section 1.5).

Postoperative outcomes vary depending on the surgical indication. For instance, THR due to all-type inflammatory arthritis is associated with a three-fold increase in 90-day mortality risk but has similar revision risks in comparison to osteoarthritic patients (294). Specifically, patients with rheumatoid arthritis are more prone to postoperative complications including revision, dislocation, and infection (295) whilst THR due to neck of femur fractures are statistically associated with increased risks of major adverse events, mortality, longer hospitalisation and intensive care unit (ICU) admission (296, 297).

There is a scarcity of evidence with regards to postoperative THR outcomes due to MBD and DDH. In comparison to primary OA, THR due to DDH was shown to be associated with a significantly longer operative time, length of hospitalisation, and need for blood

transfusions, as well as a greater risk of complications, readmissions and re-operations, albeit not significant (298).

1.9.2.2. Surgeon and unit

Several studies attempted to determine the effect of surgeon and hospital volume on outcomes following THR (291, 299-303). However, the variations amongst healthcare institutions, training programmes, and the lack of a uniform volumetric threshold of what constitutes a low versus high volume, are all factors that may account for the observed variations and hinder the generalisability of findings.

A meta-analysis of forty-two retrospective and two prospective studies from twelve countries demonstrated superior outcomes following THR in high volume hospitals as represented by statistically significant lower rates of surgical site infections, shorter length of stay, and a lower 30-day to 1-year mortality risk and 90-day complications (291). However, the lack of volume-thresholds and variations of effect size across studies hampers the external validity of the findings.

Similar to other studies conducted in the United States (304-306), a recent nationwide study using an inpatient American database of over five million patients undergoing primary elective unilateral total hip or knee replacement between 2002 and 2011 found lower risk-adjusted mortality and incidences of major complications following THR in high-volume hospitals (303). Using the American Joint Replacement Registry data between 2012 and 2020, Oakley *et al.'s* observation study on 4,447 patients showed no significant differences in one-year postoperative functional outcomes, as measured by the Hip Disability and Osteoarthritis Outcome Score (HOOS), in terms of surgeon volume ($p = 0.11$) and unit volume ($p = 0.22$). However, failing to account for baseline characteristics may have introduced bias to the outcomes (302). Once more, the variations in healthcare

infrastructure in the United States, hospital settings and surgical practice impose constraints on the broad applicability of results.

With regards to surgeon volume, a Canadian study using the Ontario health administrative database of over 35,000 patients who received primary elective THR between 2002 and 2009 showed that patients of surgeons who performed less than thirty-five procedures in the year preceding index procedure had increased risks of dislocation and revision (307). The absence of baseline data control for surgical approach, implant and fixation types, coupled with the limitations of a regional database, constrain the robustness of conclusions.

Surgeon experience was studied in an observational SHAR study that included ten hospitals in western Sweden and showed similar outcomes in terms of one-year postoperative functional improvements, pain relief and patient satisfaction following THR if performed by a speciality-certified orthopaedic surgeon, irrespective of their years in practice (292). Sayers *et al.* further analysed the effect of surgical volume and THR revision risk using the NJR for England and Wales differentiating *between-consultant* and *within-consultant* effects (308). Between-consultant was defined as a cross-sectional comparison between the surgical volume of one consultant and another in the year prior to the index procedure, and is confounded by unit-level factors, whilst within-consultant was based on time-series analyses of an individual consultant comparing volume change across time. Whilst the authors reported a strong near linear volume-revision reduction in the revision risk *between consultants*, the lack of association *within consultants* mean that changes to surgeon volume alone are unlikely to have a significant impact on THR revision risk.

1.9.2.3. Surgical approach

Whilst each surgical approach for primary THR is associated with a significant learning curve, there is moderate-to-strong evidence to suggest comparable outcomes among the most used surgical approaches (for more details see Section 1.5). These include the posterior approach (PA), direct anterior approach (DAA), direct lateral approach (DLA), minimally invasive (MIS-PA, MIS-DAA, and MIS-DLA), and supracapsular percutaneously assisted THR.

A high-quality meta-analysis of sixty-three RCTs and 4,859 patients showed no statistically significant differences in terms of hip score, QoL, pain score, length of stay, bloods loss, and cup abduction and anteversion angles among all surgical approaches when compared to PA, however, there was a longer operative time, with the exception of DLA which had smaller improvements in hip score (309). Although the authors rated the certainty of evidence based on risk of bias assessment, imprecision, heterogeneity and publication bias, the intrinsic variations in implant selection and in surgical expertise and volume may have skewed the results.

Similarly, Koster *et al.'s* meta-analysis of eight RCTs and three prospective observational studies compared different surgical approaches for THR to dislocation risk and again showed no statistical significance among all approaches (310). It is noteworthy that the limited number of RCTs assessing the dislocation risk for DAA used in the study prompt the need for further research in the field. In contrast, a population-level retrospective cohort study of 5,986 propensity-score matched patients who underwent THR in Ontario, Canada, found a small but statistically significant increased risk of one-year major surgical complications for DAA (HR 2.07; 95% CI 1.48-2.88) when compared to PA or DLA; infection (HR 3.27), dislocation (HR 2.63) and revision (HR 1.75) (311).

Blom *et al.* used the UK NJR to investigate the effect of surgical approach on revision, mortality and PROMs outcomes following primary elective THR (293). Though the authors showed no clinically significant differences in terms of PROMs among groups, lateral approaches were associated with worse death and revision outcomes when compared to PA. There remains a need for prospective clinical trials to investigate the effect of surgical approach on THR outcomes whilst adjusting for patient characteristics, surgeon factors (learning curve, experience and volume), and implant choice.

1.9.3. Implant factors

In THR, implant factors tend to have a higher modifiable influence on outcomes than patient or surgical factors. In the United Kingdom, the "Getting It Right First Time" (GIRFT) programme aimed to minimise the unwarranted variations in outcomes across the publicly-funded National Health Service (NHS), particularly in elective orthopaedic services, prompting the need to learn from 'optimal' performance (312, 313). An observational NJR study highlighted that implant choice alone, rather than surgical or unit factors, is responsible for achieving superior outcomes following primary elective THR (218). In the United States, on the other hand, where patients and healthcare insurance programmes are primarily responsible for covering healthcare-associated expenses, cost appears to be a key factor driving implant choice (314, 315).

THR implant design principal differences include: modular (unconstrained) versus monoblock (constrained) components; articulation surfaces (such as MoM, MoP, ceramic); components size; resurfacing versus THR; and fixation type (cemented, cementless or hybrid). The latter forms a substantial part of this thesis and is elaborated on in the following two sub-sections.

Modular THR implants were designed to correct anatomical reconstruction and allow intraoperative flexibility in adjusting the implant's position, angle, offset and size. Limited studies assessed the effect of modularity alone on postoperative outcomes. Whilst modular implants were reported to have some increase in risks of wear, osteolysis, and implant failure compared to monoblock, the literature is poor and prospective trials are needed (316-318).

Articulation surfaces are broadly divided into four main categories based on materials used for femoral head and acetabular bearing surface: metal-on-metal (MoM), metal-on-polyethylene (MoP), ceramic-on-polyethylene (CoP), and ceramic-on-ceramic (CoC). Observational registry studies suggest that MoM implants are associated with a higher failure rate when compared to MoP (319). However, the choice of implant recorded in large datasets is likely to due to patient characteristics introducing bias to between-implant comparisons. A systematic review and a meta-analysis of RCTs both reported no difference in survivorship, clinical outcomes and PROMs between MoM and other bearing surfaces (320, 321). However, MoM was associated with a statistically significant higher levels of ions whilst bearings containing conventional polyethylene had worse outcomes in terms of implant performance, PROMs and wear levels. On the other hand, registry data suggests that the use of CoP bearings improve implant survival, when compared to MoP (170, 228), a finding that was echoed in the previously mentioned meta-analysis (320, 321). A network meta-analysis of RCTs using implant combination selection in THR found that [MoM + small head size + cemented fixation] combination and HRA increased revision risk when compared to [MoP + small head + cemented fixation] combination (322). CoP combinations, on the other hand, had superior functional HHS outcomes with no difference in outcomes regarding the use of conventional polyethylene materials. However, the lack of long-term survivorship data in RCTs, and the limitations related to

data heterogeneity and calculated treatment effects, as well as the inability to control for patient characteristics using meta-regression models may all limit the certainty in findings.

Femoral head sizes in THR typically range from 22.25 mm to 50 mm in diameter. Similar to modularity, only a few RCTs studies compared the effect of different head sizes of varied bearing surfaces on dislocation outcomes after THR, primarily due to the large number of patients needed to detect a statistically significant difference of a rare occurrence. Four out of five RCTs reported no significant difference in revision risk based on head size (320, 323). Howe *et al.'s* powered RCT found a significant difference in dislocation incidence between 36 mm and 26 mm MoP at one-year following primary THR, thus favouring 36 mm heads (324). The use of larger sizes has increased in recent years with several arthroplasty registries reporting that 32 mm and 36 mm femoral head sizes are now most commonly used (325). Larger head sizes allow for a trade-off between increased stability and decreased THR survivorship. Based on the available evidence, the below conclusions can be drawn:

- ≤ 32 mm femoral head sizes appear to have increased risk of dislocation and lower volumetric wear and frictional torque in MoP;
- 32 mm MoP heads tend to improve implant survivability whilst CoP seem to be safer for larger implants; and
- > 36 mm head sizes: there is no sufficient evidence to support superior functional or range of motion benefits, however, ≥ 36 mm heads have a lower risk of revision due to dislocation (325).

It is argued, however, that restoring the anatomical femoral head size (~ 55 mm in males and ~ 48 mm in females), commonly used for HRA, achieve better functional outcomes

(326, 327). In terms of hip resurfacing arthroplasty, the debate is ongoing regarding postoperative outcomes compared to THR. Observational studies using a variety of statistical methods, such as propensity score matching and survival analysis, and adjustments for patient characteristics and confounding factors, report lower mortality risk for HRA in comparison to THR (12, 13, 328). Despite statistically significant adjusted differences between resurfacing and other THR implants for mortality outcome ($p = 0.012$), some authors have attributed this decreased risk to patient selection bias, rather than causal effect, where younger active patients receive the intervention (12, 14).

On the other hand, the AOANJRR and NJR for England and Wales report higher rates of revision for HRA in comparison to THR (6, 7). Observational studies using NJR reports similar revision risks for MoM HRA in males but significantly reduced implant survivorship in females, likely due to greater risks of osteoporosis and metal debris (319). This is in agreement with a network meta-analysis of fifteen RCTs, which found statistically significant increased revision risk for HRA compared to [MoP + small head + cemented fixation] combination (HR 10.38; 95% CI 2.32 - 49.40) (322). In contrast, a recent meta-analysis of eighteen RCTs comparing HRA to THR mostly found no statistically significant differences in the risks of revision, complications, and metal ion levels, however, HRA was significantly associated with less blood loss and longer operative time (329). The limitations of using revision as an endpoint outcome when making comparisons between implants include different revision thresholds, ease of revision and censoring observation bias (see Section 1.8 for more details).

There is a misconstrual of evidence regarding functional outcomes following HRA in comparison to THR. Whilst some studies reported no significant functional advantages of HRA over THR, they often failed to measure activity-specific measures (322, 329-331). A cohort of patients from the SHAR who had undergone MoM HRA and conventional THR

was randomly selected for one-to-one matching. These patients were then requested to fill out PROM questionnaires at a seven-year follow-up interval (332). The study found superior hip function outcomes, particularly in the function of sport and recreation domains. This is in agreement with other studies reporting superior outcomes for HRA including: better functional outcomes; earlier return to sport; restoration of native hip biomechanics; decreased proximal femoral stress shielding; and recreating patients' normal gait and range of movement (133-135, 333-335). Therefore, high-quality RCTs are required to evaluate functional and activity-specific measures for HRA.

1.9.4. Summary

While an extensive body of literature delineates patient-, surgical- and implant-specific factors contributing to different outcomes following THR, the increasing complexity of data and the interplay among these factors require the application of advanced analytical methods. The application of machine learning (ML) techniques may offer a different perspective and enhance our understanding of this intricate landscape. Chapters 3, 5 and 6 apply ML models and compare their predictive ability to logistic regression methods whilst providing insights into the factors driving these predictions.

Most conclusions made regarding patient factors are often taken from survival studies examining implant performance and are not particularly designed to answer these questions. For instance, delving into the potential associations between varying BMI classes and surgical indications and their implications for mortality and revision risks raises the question of whether predictive models may unravel intricate patterns. Chapter 4 uses ML-survival models to identify determinants of death and revision post THR.

Whilst limited modifications can be made to patient factors in the short-term, there is still the question of whether there are any implant-specific factors that may allow us to

optimise postoperative outcomes. Chapter 7 uses propensity-score matching techniques to compare outcomes between cemented and cementless THR (Sections 1.10 and 1.11 discuss the effect of different fixation methods in THR). Chapters 8 and 9 use a biomechanical approach to understand bone strain behaviour during femoral cementless broaching and seating, in an attempt to minimise fracture risk and enhance the fixation.

1.10. Fixation in joint replacement

Aseptic loosening is the leading cause for implant failure in joint replacements and in THR, accounting for 13.2% and 24.6% of first-time revisions in the SHAR and NJR for England and Wales, respectively (5, 6). Therefore, achieving a long-term stable bone-implant fixation continues to present a formidable challenge for surgeons and implant designers.

1.10.1. Cemented fixation

Themistocles Glück was among the first surgeons to introduce the concept of cementation in arthroplasty in the 1880s, however, its construct is credited to Sir John Charnley using polymethylmethacrylate (PMMA) bone cement in the 1960s (336).

The use of cemented fixation in THR varies across countries, with uncemented fixation now becoming the primary method used in the majority of countries. In the United Kingdom, cemented fixation halved between 2006 (40.3%) and 2021 (21.8%), whilst all-cementless and hybrid (cemented femoral stem and uncemented cup) fixation methods increased by nearly 2.5 times, and were the most common method of fixation between 2021 and 2022 (cementless^{2021} 35.4%; hybrid^{2021} 38.1%) (6). In Sweden, the use of all-cemented fixation in THR declined from 90% in 2000 to 50% in 2020, however, it remained the most common method of fixation in 2021. On the other hand, all-cementless fixation

witnessed a steady increase during the previous two decades (2000: 2.4%; 2020: 32%), whilst hybrid and reverse hybrid (cementless stem, cemented cup) fixations formed equal portions of the remaining percentage (8% each) with no HRAs reported in 2021 (5). In Australia, on the other hand, figures from 2021 show that well over 95% of THR are implanted using cementless (61.6%) or hybrid (36.3%) fixations, with a steady decline in the use of cement since 2003 (~15% in 2003 to 2.1% in 2021) (7).

1.10.2. Cementless fixation

Prior to the widespread adoption of cement and Charnley's concept of low friction arthroplasty, uncemented implants were the prevailing choice in joint replacements. In the late 1950s, McKee and Watson-Farrar used an all-cementless implant on forty patients with hip disability and reported 51% fair clinic results (337). On the other hand, Charnley's cemented THR showed mean survival of 84.7% and 74.3% at 15-year and 20-year postoperatively, respectively (338). However, the outcomes of other prosthesis were markedly poor, primarily due to loosening that was observed in over a third of the cases (339), which was at that time attributed to the use of cement, then termed "cement disease" (340). Other concerns of cemented implants included: cement properties prone to lysis and destruction (341); the high prevalence of macrophages and giant cells in post-mortem studies (106); and increased mortality risk associated with cement use (1). This led to the resurgence of cementless fixation in the early 1970s with Muller devising the "Self-Locking" anchorage in 1977 (342).

Two principal mechanisms support anchoring cementless implants to the bone: the interlock system and osseointegration. Primary stability is manifested through the press-fit of the prosthesis into the bone whilst long-term fixation is facilitated using a surface coating to promote bone-implant adherence, also known as osseointegration. Historically, cementless implants had a smooth surface solely relying on the self-locking system to

achieve primary stability and was associated with poor implant survival outcomes (343). Advancements in implant engineering allowed the development of coating materials that promote bone *ingrowth* into a porous surface of approximately 100 – 200 μm for maximal growth, or *ongrowth* onto an outer coarse surface (340). In addition to these coatings, bioactive materials such as hydroxyapatite (HA) and tricalcium phosphate (TCP) are used to promote osteoconductivity, in other words, to support bone growth and formation through stimulating cell adhesion, and the proliferation and formation of the bone extracellular matrix (344).

1.11. Fixation in Total Hip Replacement

The evidence regarding the effect of fixation on postoperative THR outcomes is contentious and appears to favour cementless fixation for superior safety whilst cemented implants tend to have lower risks of revision and periprosthetic fracture.

A nationwide observational registry study using the SHAR on 170,413 patients using KM survival analyses between 1992 and 2007 identified that uncemented implants for THR are inferior in terms of survivability in comparison to cemented implants (345). This was primarily due to the loosening of uncemented *cups* but the overall better performance of cementless *femoral* stems, which had a lower risk of aseptic loosening but an eight-fold increased risk of occult periprosthetic fractures in the first two years postoperatively. However, cementless implants formed only 8,953 (5.54%) of the overall analyses and the group was heterogenous with ten implant designs accounting for over 90% of all uncemented implants. In the United States, Chen *et al.'s* study of twenty-eight states included 447,902 eligible patients (92.14% cementless) and showed a higher rate of periprosthetic fractures in cementless implants but lower risks of death and readmission when compared to cemented fixation during the years 2016 and 2017 (346). Once again,

the significant discrepancy between groups coupled with the inclusion of only 60% of the US population hinder the internal and external validity of findings. Zhang *et al.* reviewed evidence from five international joint registries that have more than a five-year follow up (Australia, England and Wales, New Zealand, Norway, and Sweden) as well as available RCTs at the time, and concluded that cemented fixation survived better in the elderly population whilst uncemented fixation is associated with an overall better long-term survivorship in younger patients (347). However, functional and QoL measures were not considered. Furthermore, a study matched 178,784 SHAR individuals who underwent either cemented or cementless fixation for total hip arthroplasty (146,818 cemented) with 862,294 unexposed individuals from the general population using Statistics Sweden, based on age, gender, and residence, and evaluated the 90-day mortality risk following THR (348). The study found a statistically significant increased adjusted risk of death up to fourteen days for the cemented cohort compared with controls (adjusted HR 1.3, 95% CI 1.11 to 1.44), but not in the cementless-control groups (adjusted HR 0.4, 95% CI 0.09 to 1.65). Similarly, an increased risk of death was found in the hybrid group (cemented femoral component), but without reaching statistical significance, leading the authors to propose that cemented femoral, but not acetabular, fixation is likely associated with higher incidences of thromboembolic events.

Corten *et al.'s* RCT evaluated long-term re-operation outcomes following cemented or uncemented primary elective THR (349). The authors reported for the remaining ninety-three patients (original sample 250, loss to follow-up = 157) superior implant survivorship for cementless THR (n = 52/93) from the ten-year follow-up term (p = 0.020). On the other hand, six other RCTs reported no statistically significant differences in short-term outcomes, up to five years, between the two fixation groups (320). A network meta-analysis of fifteen RCTs using [MoP + small head size + cemented fixation] implant

combination as a reference group identified cemented fixation in [MoM + small head size + cemented fixation] combination to have increased short- and long-term revision risks but no statistically significant difference for cementless implant groupings (322).

Whilst the debate continues regarding the effect of fixation methods on postoperative outcomes, it is clear that the results are multifactorial encompassing unmeasured variables, such as surgical technique and experience, and require a personalised decision-making approach to improve patients' outcomes (350).

1.12. Machine learning and predictive modelling

Machine learning (ML) techniques can effectively handle complex and non-linear data patterns whilst minimising confounding factors related to selection bias, which are challenging for traditional statistical models such as logistic regression (9). In terms of predictive accuracy, several studies have demonstrated the superior power of ML in healthcare over traditional risk stratification approaches (351-356). Such models can scale horizontally across varied computing environments, preventing performance degradation across iterative cycles.

Unlike traditional statistics, ML algorithms do not rely on strict assumptions or hypotheses (357). They possess the capability to perform feature engineering by automatically identifying relevant features from high-dimensional datasets and uncovering hidden patterns. This approach provides less biased insights into the underlying drivers of the raw data (357). The implementation of ML models offers the advantage of providing adaptable and personalised recommendations, making them suitable for the dynamic data distributions and implant advancements in hip arthroplasty (358).

1.13. Declaration of interests

The focus of this thesis is hip replacement outcomes with an engineering focus on cementless short stem femoral implant. Studies on joint replacement are often supported by the manufacturers of these implants. The department where this research was conducted receives partial funding from industry grants, and charitable funds from the Michael Uren Foundation. The principal supervisor of this thesis has patents and stocks (and stock options) for Embody Orthopaedic Limited, and previously Additive Instruments Limited. The co-supervisor is from an engineering background and has multiple crossovers with the industry was part of the design team for the two companies. The latter, Additive Instruments Limited, was recently acquired by Smith & Nephew Limited.

The author of this thesis was enrolled as a PhD student at the Department of Surgery and Cancer at Imperial College London and received a fully funded President's PhD scholarship, which includes tuition fees, monthly stipends and modest research consumables. Research consumables beyond that was provided by the charitable funds available to the principal supervisor. JRI provided the instruments and implants used during the biomechanical testing for Part II of this thesis. None of the funding bodies had any input into the design, development, analysis or write up of this work nor any control over the decision to disseminate the findings in the form of presentations or publications.

1.14. Scope and structure of thesis

The aim of this thesis was initially to evaluate the binary discriminative ability of different interpretable machine learning models in predicting post-operative outcomes in primary elective THR, and to delineate the patient, surgical, and implant-design factors that

improve health-related quality of life measures and reduce failures (re-operation, revision, periprosthetic femoral fracture necessitating re-do procedures, and mortality). Findings from these evaluations shaped the efforts to provide practical applications to minimise periprosthetic fracture risks associated with cementless femoral implants, providing bidirectional research translation from the bedside back to the bench. By understanding both aspects, this thesis aim to enhance the predictability and outcomes of THR procedures. The structure of the thesis is outlined below.

In Chapter 3, interpretable predictive models for mortality and implant failure risk are evaluated using two national joint registries: the Swedish Hip Arthroplasty Register (SHAR), and the National Joint Registry for England and Wales. This represents the largest and most comprehensive analysis to date. Multiple predictor factors are used to train the algorithms and the outcomes were evaluated at various endpoints ranging between thirty days and five years post-operatively. Insights from these models, whilst clinically relevant, are not subject to inferential testing assumptions. Therefore, Chapter 4 will seek identify the determinants predisposing to poor outcomes using machine learning-selected variables, which applies a penalty to the coefficients and shrinks others towards zero, and Cox proportional hazard regression survival models, followed by multivariate analyses of these features.

Chapter 5 concerns the forecasting of meaningful short-term improvements in the health-related quality of life measures after primary THR using the SHAR. Similar to Chapter 3, the key features behind the internal decision making for the top performing machine learning algorithms are highlighted.

In view of the findings in Section 1.11 and the findings in Chapters 3 and 4, high revision rates, particularly in the short-term, appear to be associated with cementless implants.

Chapters 6 and 7 concerns delineating a specific reason for implant failure, periprosthetic femoral fracture (PPFF), in the SHAR. Like Chapter 5, Chapter 6 will develop machine learning models that forecast the risk of PPFF necessitating a re-do procedure and identify the predictors of femoral fractures from a machine learning perspective. Chapter 7 will examine the effect of cementless fixation on mortality, revision and periprosthetic femoral fracture using the SHAR. Patients receiving each fixation method for primary THR are closely matched using propensity score techniques to minimise confounding by indication.

Building on the findings gained from the evaluations in Part I, the two major studies in the second part of this thesis use a biomechanical lens to understand the optimal surgical approach to seat cementless femoral stems to reduce the risk of periprosthetic fracture whilst ensuring sufficient bone-stem fixation. Part II investigates the effect of three different energy levels on bone strain behaviour, femoral fracture risk, and implant fixation to the bone during cementless short-stem femoral broaching and implant seating. Chapter 8 will use a digital image correlation technique and a validated rig that accounts for soft tissue compliance to measure strain behaviour in the proximal femur using synthetic bones of low and normal mineral densities. Chapter 9 will further test these findings in cadaveric "healthy bone" specimens using rosette strain gauges, providing guidance for further bench-side experiments in the field.

Chapter 2 Data and machine learning methods

2.1. Introduction

The research studies included in this thesis use several data sources. This chapter describes the datasets, initial processing and cleaning of datasets, classification of variables and the generic machine learning methods used. Each subsequent chapter has a specific methodology and the statistical inference and machine learning methods for the analyses are described. The biomechanical studies (Chapters 8 and 9) build on emerging findings from the machine learning studies. Throughout Chapters 3 to 7, all data cleaning, statistical and machine learning analyses are performed using R statistical computing environment version 4.3.0 (R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/), unless stated otherwise.

2.2. The Swedish Arthroplasty Register

2.2.1. Background

The Swedish Hip Arthroplasty Register (SHAR), founded in 1979, was the world's first national hip register collecting data on primary hip arthroplasty to provide information on quality and implant safety. The Swedish Arthroplasty Register (SAR) was formed in 2021, replacing SHAR by merging SHAR with the Swedish Knee Arthroplasty Register, including hemiarthroplasties and knee osteotomies. SAR describes its aim as being "to maintain a continuous follow-up of the health care quality, contribute to improvement work and research to gain knowledge and give the patients the best possible care"(169).

The SAR now contains data from over half a million primary hip replacements as well as approximately 92,000 hip arthroplasty re-operations (5). Data are reported in several ways and primarily communicated to the public in the form of an annual report. The data are also reported online with filtering options (https://slr.registercentrum.se/) and are available to healthcare professionals and provide implant suppliers with special access to reports on inserted implants (https://krh.registercentrum.se/). The register has extensive research involvement and collaborates with researchers around the word; requests for accessing SAR data for the purposes of research are considered by the SAR steering group.

2.2.2. Data collection

The SAR collects data on patients' demographics, arthroplasty procedures and patient reported outcome measures (PROMs). Data are collected by several health care professionals using standardised forms as well as patients using a PROM-questionnaire and are entered into the register web portal at a county/unit level. Data related to primary and re-operation hip arthroplasty are collected for all nationwide public and private hospitals in Sweden and report to the register with an estimate completeness of 98%(169).

2.2.3. Data extract

Release of data for the purposes of this study was approved by the SHAR Steering group on 12th December 2019 (Appendix 1, A1.1). One extract was produced containing all records of primary hip replacement, hip hemiarthroplasty, re-operation hip replacement with linked mortality, re-operation, and PROM data. The extract contained patients' demographics and operative and component-level data that allowed for the detailed analyses of patient, surgical, and implant factors. Given the low level of data completeness prior to the year 2000, the extract covered the period between January 2000 and August 2018.

2.2.4. Data cleaning and initial processing

The main extract contained records for 442,546 hips. Prior to conducting data cleansing, numerous variables within the dataset exhibited missing values, indicating the presence of empty or null values. Data were imported into R Statistical Programming and coded. All cases performed for patients below the age of 18 years old after data cleaning (131 observations), and with a diagnosis of trauma, tumour, and non-osteoarthritic degenerative conditions were removed.

2.2.4.1. Data cleaning

All variables were assessed for missing data. Body Mass Index (BMI) or American Society of Anaesthesiologists (ASA) Classification had a low level of completeness before year 2008 (>50% missing data). In order to mitigate the impact of extensive missing data, the choice was made to exclude the records with missing values. After exclusion of cases where the register lacks data for certain variables, the cleaned dataset contained 154,595 primary hip arthroplasty procedures. The statistical summary for the processed data is shown in table 2-1.

Swedish Hip Arthroplasty Register	
Number of operations	154,595
Age at surgery, mean (SD)	67.9 (10.6)
Sex (Male)	66,868 (43.3%)
Body Mass Index (BMI)	
Underweight	1,191 (0.8%)
Normal	48,037 (31.1%)
Overweight	66,408 (43.0%)
Obesity Class I	29,757 (19.2%)
Obesity Class II	7,647 (4.9%)
Obesity Class III	1,555 (1.0%)
ASA grade	
1	36,826 (23.8%)
2	92,076 (59.6%)
3 +	25,693 (16.6%)

Side (Right)	70,000 (45.3%)
Unit type	
University Hospital	12,244 (7.9%)
County Hospital	47,052 (30.4%)
Rural Hospital	60,254 (39.0%)
Private Hospital	35,045 (22.7%)
Pre-operative Diagnosis	
Primary arthrosis	141,412 (91.5%)
Inflammatory	1,962 (1.3%)
Childhood dx	3,256 (2.1%)
Idiopathic necrosis	3,741 (2.4%)
Secondary arthrosis	4,097 (2.7%)
Other	127 (0.1%)
Incision type	
Posterior	83,146 (53.8%)
Direct lateral	70,924 (45.9%)
Direct anterior	324 (0.2%)
Trochanteric	173 (0.1%)
Other	12 (0.0%)
Prosthesis group	
Cemented	95,490 (61.8%)
Cementless	33,479 (21.7%)
Hybrid	5,030 (3.3%)
Reverse hybrid	19,528 (12.6%)
Resurfacing	1,068 (0.7%)
Cement type	
High viscosity with antibiotic	55,195 (35.7%)
Low viscosity with antibiotic	876 (0.6%)
High viscosity without antibiotic	58,915 (38.1%)
Low viscosity without antibiotic	11 (0.0%)
Cementless, hybrid or resurfacing	39,577 (25.6%)
Cup fixation	
Cemented	115,020 (74.4%)
Cementless	39,575 (25.6%)
Cup resurfacing	1,528 (1.0%)
Cup modularity (Monoblock)	115,818 (74.9%)
Cup articulation	
Metal (standard)	235 (0.2%)
Metal (resurfacing)	1,531 (1.0%)

Ceramic	637 (0.4%)
Dual mobility (monoblock)	1,278 (0.8%)
Dual mobility (modular)	5 (0.0%)
Poly (standard)	44,116 (28.5%)
Poly (x-link)	106,628 (69.0%)
Unclear	165 (0.1%)
Cup modification	
Standard	100,836 (65.2%)
Liped	12,319 (8.0%)
Dual articular	1,146 (0.7%)
Constrained	10 (0.0%)
Unclear	40,284 (26.1%)
Cup inner diameter, mean (SD)	31.49 (3.05)
Stem fixation	
Cemented	101,585 (65.7%)
Cementless	53,010 (34.3%)
Stem resurfacing	1,068 (0.7%)
Stem modularity (Modular)	153,466 (99.3%)
Stem articulation	
Metal	127,730 (82.6%)
Ceramic	25,766 (16.7%)
Unclear	1,099 (0.7%)
Femoral head size, mean (SD)	31.49 (3.03)
Outcomes	
Death	15,177 (9.8%)
Re-operation	7,049 (4.5%)
Revision	5,179 (3.3%)

Table 2-1 Baseline characteristics of the SHAR extract after data cleaning without PROMs

2.2.4.2. Data Classification

In total, the structured dataset included twenty-one variables: six patient variables (age, gender, BMI, ASA, side of operation, and the type of hospital), four surgical variables (diagnosis group, incision type, prosthesis group, and cement type), and eleven implant-specific variables which were each subdivided into six cup variables and five femoral variables (fixation, resurfacing, articulation, modular or monoblock, and implant size; in

addition to the sixth cup-related variable "modification"). Each variable is described and summarised in more detail in Appendix 2.

2.2.4.3. Linkage

The SHAR dataset provides information on mortality, causes, and time to event (re-operation, revision and/or death). The Tax Office maintains a record of death dates, which are periodically synchronised with the SHAR and other government databases. Therefore, linking SHAR to national statistics data was not required.

2.3. The National Joint Registry for England and Wales

2.3.1. Background

The establishment of the National Joint Registry for England and Wales (NJR) took place in 2002, and data collection commenced the following year. The stated mission of the organisation is "to collect and analyse high quality and relevant data about joint replacement surgery in order to provide timely warnings of issues relating to patient safety" (170).

The NJR now is the largest orthopaedic registry in the world with over 3.7 million joint replacement procedure records submitted (170). Similar to the SHAR, data are disseminated through multiple channels. The primary mode of reporting NJR data is through an annual clinical report. NJR feedback services provide access to different stakeholders including *implant manufacturers* to access component-level data (supplier feedback), *surgeons* to access their own real-time data online (clinician feedback), and *consultant-level and senior management reports* to provide information on lead surgeons and price benchmarking figures (consultant feedback and management feedback) (359). For

research purposes, data can be made available upon request, and such requests are evaluated by the NJR research sub-committee.

2.3.2. Data collection

Standard forms are used to collect data for primary and revision hip replacements with separate consent forms provided by the patient pre-operatively. Data entry onto the NJR database is done via an online platform by staff at each individual NHS and private hospital in England and Wales, and since 2013, Northern Ireland has been added. Cases from England, Wales, and Northern Ireland are included.

2.3.3. Data extract

The NJR Research Sub-committee declined to release the data for the purposes of this study in 2019. A collaboration with the University of Sheffield, and the NJR research team, was successfully established in 2021. The author of this thesis worked closely with the research team at Sheffield to externally validate the SHAR-developed machine learning algorithms. As per the collaborating team, a single combined extract containing all records of primary hip replacements with linked mortality and revision data as well as component-level data were released to them by the NJR. Data extracted were from the start of 2009 through to August 2018.

2.3.4. Data cleaning and initial processing

The main extract contained records for 1,141,014 hips. Data cleaning and initial processing followed the same methodological steps utilised for the SHAR extract. Prior to conducting data cleansing, numerous variables within the dataset exhibited missing values, indicating the presence of empty or null values. Data were imported into Python's Statistics Programming and coded. All cases performed for patients below the age of 18 years old,

and with a diagnosis of trauma, tumour, and non-osteoarthritic degenerative conditions were removed.

2.3.4.1. Data cleaning

Similar to the SHAR, all variables were assessed for missing data. Body Mass Index (BMI) or American Society of Anaesthesiologists (ASA) Classification had a low level of completeness before year 2008 (>50% missing data). To mitigate the impact of extensive missing data and replicate the methodology used for the SHAR data extract, missing values records were removed. After exclusion of cases, the cleaned dataset contained 355,933 primary hip arthroplasty procedures. The statistical summary for the processed data is shown in table 2-2.

National Joint Registry	
Number of operations	355,933
Sex (Male)	141,677 (39.8%)
Body Mass Index (BMI)	
Underweight	2,561 (0.7%)
Normal	68,072 (19.1%)
Overweight	139,731 (39.2%)
Obesity Class I	95,078 (26.7%)
Obesity Class II	37,332 (10.5%)
Obesity Class III	13,159 (3.7%)
ASA grade	
1	45,247 (12.7%)
2	247,372 (69.5%)
3 +	63,314 (17.8%)
Side (Right)	197,009 (55.3%)
Unit type	
University Hospital	56,863 (16%)
Public Hospital	221,759 (62.3%)
Private Hospital	77,311 (21.7%)
Pre-operative Diagnosis	
Primary arthrosis	335,665 (94.3%)
Inflammatory	3,506 (1.0%)
Childhood dx	5,288 (1.5%)

Idiopathic necrosis	8,455 (2.4%)
Secondary arthrosis	416 (0.1%)
Other	2,603 (0.7%)
Incision type	
Posterior	243,953 (68.5%)
Direct lateral	100,284 (28.2%)
Direct anterior	111 (0.0%)
Trochanteric	517 (0.2%)
Other	11,068 (3.1%)
Prosthesis group	
Cemented	116,130 (32.6%)
Cementless	141,901 (39.9%)
Hybrid	85,877 (24.1%)
Reverse hybrid	11,975 (3.4%)
Resurfacing	50 (0.0%)
Cement type	
High viscosity with antibiotic	209,686 (58.9%)
Low viscosity with antibiotic	1,320 (0.4%)
High viscosity without antibiotic	3,026 (0.9%)
Cementless	141,901 (39.8%)
Cup fixation	
Cemented	128,105 (36%)
Cementless	227,828 (64%)
Cup resurfacing	50 (0.0%)
Cup modularity (Monoblock)	
Monoblock	127,173 (35.7%)
Other	228,760 (64.3%)
Cup articulation	
Metal	2,165 (0.6%)
Ceramic	50,673 (14.2%)
Dual mobility	1,677 (0.5%)
Poly (standard)	169,136 (47.5%)
Poly (x-link)	132,232 (37.2%)
Stem fixation	
Cemented	202,043 (56.8%)
Cementless	153,890 (43.2%)
Stem articulation	
Metal	221,640 (62.3%)
Ceramic	134,293 (37.7%)

Outcomes	
Death	26,395 (7.4%)
Revision	5,195 (1.5%)

Table 2-2 Baseline characteristics of the NJR extract after data cleaning

2.3.4.2. Data Classification

Like the SHAR data classification, the structured dataset included twenty-one variables: six patient variables (age, gender, BMI, ASA, operation side, and unit type), four surgical variables (diagnosis group, incision type, prosthesis group, and cement type), and eleven implant-specific variables which were each subdivided into six cup variables and five femoral variables (fixation, resurfacing, articulation, modular or monoblock, and implant size; in addition to the sixth cup-related variable "modification"). Each variable is described and summarised in more detail in Appendix 2.

2.3.4.3. Linkage

The NJR data extract was linked to the Hospital Episode Statistics (HES), which provides information on mortality. Data linkage was provided by the NJR research team with access provided to the University of Sheffield's team.

2.4. Hospital Episode Statistics

2.4.1. Background and data collection

Hospital Episode Statistics (HES), founded in 1989, is a data repository containing details about hospital admissions, outpatient appointments and emergency department attendances in the NHS in England (360). HES include episode-level data on private patients treated in NHS hospitals and care delivered by treatment centres funded by the NHS, meaning that each patient may have multiple HES records. NHS trusts are responsible for collating, coding, and incorporating data onto the HES database using patient notes and discharge summaries. HES is managed by NHS Digital with data

including demographics (age and gender), diagnosis, procedure codes, length of stay, admission status, discharge details and mortality.

2.4.2. Data extract and linkage

The NJR data extract was linked to the Hospital Episode Statistics (HES), which provides information on mortality. Data linkage was provided by the NJR research team with access provided to the University of Sheffield's team.

2.5. Safety and Implant Failure Outcome Measures

Outcomes of hip joint replacement can be divided into safety outcomes, such as mortality and morbidity rates related to the operation, and implant failure outcomes such as rates of re-operation and revision surgery (361). The key outcome measures were the occurrence of death, re-operation, revision, and periprosthetic fracture that required a revision procedure, within a number of time intervals. Machine learning predictive models for mortality and revision were developed on the SHAR dataset (80%), internally validated (20%) and then externally validated using the NJR dataset. For the occurrences of re-operation and periprosthetic fracture requiring revision surgery, the machine learning models were only internally validated using the SHAR dataset for pragmatic reasons.

2.5.1. Mortality

In both the SHAR and NJR registry datasets, the term "mortality" refers to the occurrence and recording of deaths and includes information such as the date of death within a specified population or cohort, in this instance, patients who underwent hip arthroplasty. The SHAR dataset is periodically synchronised with the Swedish Tax Office whilst the NJR dataset needed to be linked to HES to maintain a record of death dates. For mortality, we considered deaths within 30 days, 60 days, 90 days and 1 year as separate outcomes.

We limited our predictive models for mortality to the one-year period since mortality beyond that timeframe can potentially be influenced by confounding factors.

2.5.2. Revision and Re-operation

In the SHAR, re-operation is defined as any additional surgery to the hip, regardless of the actions taken to any of the implant components (replaced, extracted, or left untouched) whereas revision is defined as a re-operation where at least one component is exchanged, extracted from or added to the prosthesis. Similar definitions are used in the NJR dataset. For re-operations and revisions, we considered operations performed within the first 6 months, 1 year, 2 years, 5 years and 10 years as separate outcomes of interest.

2.5.3. Periprosthetic Fracture

Periprosthetic fracture necessitating revision surgery is a specific subset of revision outcomes within the SHAR, where the prosthesis undergoes modification by exchanging, extracting, or adding at least one component due to the occurrence of a periprosthetic fracture. In this study, our outcome variable was defined as the occurrence of a revision surgery conducted between thirty days and one-year following the primary procedure, specifically attributed to a periprosthetic fracture.

2.6. Patient Reported Outcome Measures

2.6.1. Background

Although the development of registry data predominantly revolves around the capturing and analysis of adverse events and complications arising from surgical procedures, the systematic evaluation of quality outcome data was influenced by the Darzi review of 'High Quality Care for All' in 2008 (362). The report emphasises the necessity of delivering high-quality care to all patients, and one of the initial steps towards achieving this is

through the systematic collection of Patient Reported Outcome Measures (PROMs) for commonly performed elective procedures. In the United Kingdom, PROMs for hip and knee arthroplasty commenced in April 2009 with the addition of varicose vein and groin hernia repair postoperative PROMs. The latter two phased out in 2017 in efforts to orchestrate a balance between cost-effectiveness and utility at a population-level (363).

SHAR commenced the inclusion of PROMs in 2002 as a pilot project and the programme progressively expanded to encompass all Swedish healthcare facilities performing total hip replacements. PROMs are collected in the form of self-administered protocols comprising the EQ-5D instrument, the Charnley class categorisation and visual analogue scales for pain and satisfaction with complete pre- and post-operative questionnaires (181). In this study, PROMs predictive models are developed and internally validated using the SHAR dataset. For pragmatic reasons, there was no external validation using the NJR for England and Wales.

2.6.2. Data collection

PROMs data in the SHAR are completed by patients using a self-administered ten-item questionnaire pre-operatively and at one, six, and ten years post-operatively, including Charnley's functional categories (A, B, and C) (106, 364), a visual analogue scale (VAS) for pain, and the generic EQ-5D measurement (365). The preoperative questionnaire has been adapted to a valid and reliable internet-based application to minimise the risk of systematic errors; however, pen-and-paper questionnaires are still in-use based on individual unit's and patients' preferences. Post-operative questionnaires are sent to patients via postal mail with enclosed return envelopes, followed by a single reminder after eight weeks (181).

In 2011, the SHAR adopted the addition of VAS Satisfaction, which addresses patient satisfaction with the outcome of the hip arthroplasty procedure and is obtained in the postoperative PROMs questionnaires. VAS Satisfaction was shown to be a valid PROM reflecting patients' experiences and correlated with the different EQ-5D dimensions (181).

2.6.3. Data linkage and cleaning

All the SHAR extracted records were linked to PROMs. Data since 2008, when PROMs were reported to the SHAR on a national basis, were included. Cases with missing either pre- or post-operative EQ-5D index, EQ-VAS values or Charnley classifications were removed. 82,526 hips were therefore available for analysis. The statistical summary for the processed data is shown in Table 2-3.

Swedish Hip Arthroplasty Register	
Number of operations	82,526
Age at surgery, mean (SD)	68.02 (10.2)
Sex (Male)	35,689 (43.2%)
Body Mass Index (BMI)	
Underweight	581 (0.7%)
Normal	25,818 (31.3%)
Overweight	36,070 (43.7%)
Obesity Class I	15,402 (18.7%)
Obesity Class II	3,879 (4.7%)
Obesity Class III	776 (0.9%)
ASA grade	
1	21,147 (25.6%)
2	49,531 (60.0%)
3 +	11,848 (14.4%)
Side (Right)	37,179 (45.1%)
Unit type	
University Hospital	6,267 (7.6%)
County Hospital	26,558 (32.2%)
Rural Hospital	32,789 (39.7%)
Private Hospital	16,912 (20.5%)
Pre-operative Diagnosis	

Primary arthrosis	76,458 (92.6%)
Inflammatory	1,108 (1.3%)
Childhood dx	1,664 (2.0%)
Idiopathic necrosis	1,529 (1.9%)
Secondary arthrosis	1,730 (2.1%)
Other	37 (0.0%)
Incision type	
Posterior	43,577 (52.8%)
Direct lateral	38,588 (46.8%)
Direct anterior	237 (0.3%)
Trochanteric	112 (0.1%)
Other	10 (0.0%)
Prosthesis group	
Cemented	54,258 (65.7%)
Cementless	15,048 (18.2%)
Hybrid	1,873 (2.3%)
Reverse hybrid	10,585 (12.8%)
Resurfacing	762 (0.9%)
Cement type	
High viscosity with antibiotic	31,367 (38.0%)
Low viscosity with antibiotic	740 (0.9%)
High viscosity without antibiotic	32,715 (39.6%)
Low viscosity without antibiotic	6 (0.0%)
Cementless, hybrid or resurfacing	17,683 (21.4%)
Cup fixation	
Cemented	64,843 (78.6%)
Cementless	17,683 (21.4%)
Cup resurfacing	1,034 (1.3%)
Cup modularity (Monoblock)	65,808 (79.7%)
Cup articulation	
Metal (standard)	179 (0.2%)
Metal (resurfacing)	1,034 (1.3%)
Ceramic	408 (0.5%)
Dual mobility (monoblock)	300 (0.4%)
Dual mobility (modular)	0 (0.0%)
Poly (standard)	31,364 (38.0%)
Poly (x-link)	49,153 (59.6%)
Unclear	88 (0.1%)
Cup modification	

Standard	54,275 (65.8%)
Liped	9,671 (11.7%)
Dual articular	278 (0.3%)
Constrained	4 (0.0%)
Unclear	18,298 (22.2%)
Cup inner diameter, mean (SD)	31.20 (3.30)
Stem fixation	
Cemented	56,893 (68.9%)
Cementless	25,633 (31.1%)
Stem resurfacing	762 (0.9%)
Stem modularity (Modular)	81,721 (99.0%)
Stem articulation	
Metal	69,316 (84.0%)
Ceramic	12,431 (15.1%)
Unclear	779 (0.9%)
Femoral head size, mean (SD)	31.20 (3.28)
Pre-operative Charnley Classification	
A	39,627 (48.0%)
B	10,026 (12.1%)
C	32,873 (39.8%)
Post-operative Charnley Classification	
A	39,729 (48.1%)
B	8,373 (10.1%)
C	34,424 (41.7%)
Pre-operative PROMs	
EQ-5D, median (IQR)	0.58 (0.59)
EQ-VAS, mean (SD)	56.0 (22.38)
Post-operative one-year PROMs	
EQ-5D, median (IQR)	0.80 (0.30)
EQ-VAS, median (IQR)	80 (25)
Outcomes	
EQ-5D index, SRM threshold (Improved)	54,734 (66.3%)
EQ-5D index, paired threshold (Improved)	32,890 (39.9%)
EQ-VAS (Improved)	57,064 (69.1%)
Combined (paired EQ-5D and EQ-VAS)	25,849 (31.3%)
Combined (SRM EQ-5D and EQ-VAS)	42,653 (51.7%)

Table 2-3 Baseline characteristics of the SHAR extract after data cleaning including PROMs

2.7. Personal contribution to data collection and analysis

As outlined in Section 2.2, a substantial portion of the data utilised in this study was sourced from the SHAR and the NJR for England and Wales. Access to the NJR for England and Wales was declined in 2019 and thus we obtained access to the SHAR dataset through a formal request and collaboration with the Swedish Hip Arthroplasty Register and the academic team at the University of Gothenburg. In order to validate our SHAR-based machine learning models, we collaborated with the academic team at the University of Sheffield who had access to the NJR dataset linked with HES data.

All analyses of the SHAR data were performed by the author of this thesis. External validation of our machine learning models using the NJR dataset was performed by a member of the academic team at the University of Sheffield in close collaboration with the author of this work.

Results, Part One:
Machine Learning & Outcomes following Primary Hip Arthroplasty

Chapter 3 Predicting mortality and failure outcomes in primary hip arthroplasty

3.1. Introduction

Considerable variability exists in clinical practice and implant selection for patients with hip osteoarthritis requiring arthroplasty. The relative outcomes of each, and their suitability for different patients, can be difficult to detect because of variation amongst patients, surgeons, and units. Careful evaluation of major adverse events plays a pivotal role in the shared decision-making process, with significant implications regarding the choice of optimal implants and surgical factors (1).

Outcomes of hip joint replacement can be divided into safety outcomes, such as mortality and morbidity rates related to the operation, implant failure outcomes such as rates of re-operation and revision surgeries, functional outcomes including range of motion and physical activity levels, and satisfaction and patient reported outcomes (361). Although the rates of death, re-operation, and revision following THR are low, it is expected that the incidence of primary THR may see over a 200% increase during the next decade (366), increasing substantially also the social and economic impacts of adverse outcomes. Therefore, novel tools for minimising unnecessary risks and optimising treatment outcomes are needed.

In order to predict outcomes following THR, observational datasets are usually analysed using multivariable regression models. From a clinical perspective, the lack of objective metrics of joint health and disease and of prospective outcome data for many surgical treatments, enables surgeons' selection and treatment biases to persist. This coupled with

the poor accuracy of surgeons in predicting patient outcomes (31, 367, 368) and the lack of consensus on optimal management strategies, especially for common conditions like osteoarthritis, prompt the need for large unbiased data and the use of targeted artificial intelligence (AI) systems. To this aim, machine learning (ML) predictive models offer an alternative which have the potential to better forecast future events, while accounting for a large number of predictor variables and their potentially complex and non-linear interactions (9). Several studies have demonstrated the superior predictive accuracy of ML in healthcare over traditional risk stratification approaches (351-356).

The aim of this chapter is to develop novel ML-based risk prediction models integrating patient, surgical, and implant features and compare their predictive accuracy to traditional statistical linear models, using large datasets from the Swedish Hip Arthroplasty Register and the National Joint Registry for England and Wales. Multiple outcomes will be predicted to accurately assess the effectiveness of different algorithms, including mortality, re-operation and revision events at multiple time points ranging from 30 days to 10 years. ML models for mortality, re-operation and revision are internally validated using the SHAR, and ML models for mortality and revision are also externally validated using the NJR. The subsequent chapter, Chapter 4, employs advanced ML inference testing techniques to elucidate the machine learning models and identify the factors that contribute to the outcomes. Chapter 5 evaluates the use of machine learning models in predicting Patient Reported Outcome Measures (PROMs) following primary hip arthroplasty. Chapters 6 and 7 develop machine learning models to predict revision risk due to periprosthetic fractures and investigate the effect of implant fixation on periprosthetic fracture outcomes using propensity score matching techniques.

3.2. Methods

3.2.1. Data sources

Chapter 2 provides a comprehensive description of the three data sources utilised in this study, including information on data collection, processing, and linkage. Whilst data extract from the SHAR is since 2000, NJR/HES linked data extract for the external validation are since 2009. Therefore, this study is in two parts; the first part uses the SHAR dataset to develop and internally validate the machine learning algorithms in predicting safety and implant failure outcomes (mortality, revision and re-operation), and the second uses NJR/HES linked data to externally validate the algorithms for each outcome. Table 3-1 represents the variables used in each dataset.

	SHAR		NJR		HES
Patient factors	Age		Age		Age
	Gender		Gender		Gender
	BMI		BMI		
	ASA grade		ASA grade		
	Side of operation		Side of operation		
	Hospital type		Unit type		Unit type
Surgical factors	Operation type		Operation type		Procedure type
	Diagnosis group		Diagnosis group		Length of stay
	Incision type		Incision type		
	Prothesis group		Prosthesis group		
	Cement type		Cement type		
Cup implant factors	Fixation		Fixation		
	Resurfacing		Resurfacing		
	Articulation		Articulation		
	Modular or Monoblock		Modular or Monoblock		
	Inner diameter		Inner diameter		
	Modification		Modification		
Femoral implant factors	Fixation		Fixation		
	Stem resurfacing		Stem resurfacing		
	Articulation		Articulation		
	Head size		Head size		
Outcomes	Death	30-day	Death (data linked with HES)		30-day
		60-day			60-day
		90-day			90-day
		1-year			1-year
	Revision	6-month	Revision	6-month	
		1-year		1-year	
		2-year		2-year	
		5-year		5-year	
		10-year			
	Re-operation	6-month			
		1-year			
		2-year			
		5-year			
		10-year			

Table 3-1 Sources of data

The structured dataset used in the study comprised a total of twenty-one variables. These variables can be categorised into six patient-related variables, four surgical variables, and eleven implant-specific variables. The six patient variables included in the dataset are age, gender, body mass index (BMI), American Society of Anaesthesiologists (ASA) score, the side of the operation (left or right), and the type of hospital where the procedure was performed. The four surgical variables captured in the dataset are the diagnosis group (indicating the specific condition or reason for the surgery), incision type (the type of surgical incision made), prosthesis group (categorising the type of hip prosthesis used), and cement type (the type of cement utilised in the procedure). The eleven implant-specific variables were further divided into six cup variables and five femoral variables. The cup variables included fixation type (how the cup component is secured to the bone), resurfacing (whether the cup is a total hip replacement or hip resurfacing), articulation type (the materials used for the bearing surfaces), modular or monoblock design (the structure of the implant), implant size, and the additional variable "modification" related to the cup. The femoral variables encompassed fixation type, resurfacing, articulation type, modular or monoblock design, and implant size.

3.2.2. Exposures

For both parts of the study, machine learning models were applied to forecast the outcomes of hip arthroplasty. Analyses were restricted to patients undergoing elective primary hip replacement. Patients who had undergone hip arthroplasty due to specific diagnoses such as trauma, tumour-related conditions, and degenerative conditions other than osteoarthritis were excluded. Patients below the age of 18 were excluded to minimise potential confounding factors that could arise from the different outcomes observed in paediatric patients (369-371).

3.2.3. Mortality and implant failure outcomes

Six machine learning models (described below in Section 3.2.4) were developed and compared based on their accuracy in predicting mortality (at 30, 60, 90 days and 1 year), revision and re-operation (at 6 months, 1, 2, and 5 years) following elective primary hip arthroplasty (Table 3-1).

Mortality included patients identified as having undergone primary hip arthroplasty in the SHAR data, which is periodically synchronised with the Swedish Tax Office records, and the NJR data linked with the HES database (see Section 2.3). Analysis of revisions involved patients who were identified as having undergone revision in the SHAR data and the NJR data. For pragmatic reasons, re-operations were only analysed in the SHAR data and not externally validated with the NJR for England and Wales. Predicting revision risk due to periprosthetic fracture, as reported by the operating surgeon, using the SHAR data is described in Chapter 6.

3.2.4. Model development, comparison, and internal validation

We applied several machine learning algorithms, namely random forest (RF), gradient boosting machine (GBM), penalised logistic regression (with the Lasso or Ridge penalties), and classification tree (with and without pruning), and compared them to traditional statistical model logistic regression to predict the occurrence of death, re-operation, and revision. For model development and internal validation, the SHAR derivation cohort of 154,595 observations was randomly split into a training dataset (N=123,676, 80% of the patients) and an internal validation cohort (N=30,919, 20% of the patients). The training cohort was used to train the ML models and to adjust their hyperparameters via cross-validation whereas the validation cohort was used to assess the models' performance on unseen data. For a more comprehensive understanding of each variable and a detailed summary of the SHAR dataset, please refer to Appendix 2.

For external validation, the machine learning models were applied to the NJR derivation cohort of 355,933 observations, which was again randomly split into a training dataset (N=284,747, 80% of the patients) and an external validation cohort (N=71,186, 20% of the patients).

Model performance was evaluated and compared in terms of the area under the receiver operating characteristic curve (AUC) in each dataset. For each time-defined outcome, we selected the best performing model to comprise a set of models referred to as SafeHip models. Based on patient-specific data on the twenty-one included variables, the SafeHip models can be applied to obtain individualised risk scores for each outcome.

3.2.5. Statistical analysis and feature importance

Kaplan-Meier (KM) estimates were employed to describe the rates of mortality, re-operation, and revision across various age and sex groups in both the SHAR and NJR datasets. The Kaplan-Meier method is a nonparametric statistical approach widely used to estimate survival or event probabilities over time, enabling us a comprehensive descriptive analysis of the study's outcomes (372).

The Bland-Altman method was employed to evaluate the concordance between the AUCs produced by the NJR and SHAR machine learning models. The top performing models were used to drive insights into the key predictor variables for each study outcome. The usefulness of these predictors was two-fold: first, to check the relevant clinical knowledge and expertise in driving these predictions, and second, to compute survival models using the lasso penalised Cox proportional hazard regression (CoxNet) model in Chapter 4.

All mathematical modelling was carried out using R statistical computing environment version 4.3.0 (R: A language and environment for statistical computing). R packages ‘survival’ (version 3.5.5), ‘gbm’ (version 2.1.8.1), ‘glmnet’ (version 4.1.7), ‘tree’ (version

1.0.43), 'rpart' (version 4.7.19), and 'randomForest' (version 4.7.1.1), were used for survival analysis, penalised logistic regression, classification tree, and implementation of RF, respectively. The Bland-Altman technique was applied using MATLAB software (2021b, MathWorks, Massachusetts, USA).

3.3. Results

3.3.1. Swedish Hip Arthroplasty Register

Within one year following the primary operation, re-operation was the most frequent outcome, with a total of 3,467 incidents, accounting for 2.24% of the cases (as presented in Table 3-2). During this same time frame, the number of revisions amounted to 2,287 cases (1.48%), while the number of deaths reached 1,301 (0.84%).

Outcome	Time from primary operation	Derivation cohort (N=154,595) N	%
Mortality	30 days	164	0.11
	60 days	265	0.17
	90 days	364	0.24
	1 year	1,301	0.84
Revision	6 months	1,815	1.17
	1 year	2,287	1.48
	2 years	3,011	1.95
	5 years	4,179	2.70
	10 years	5,133	3.32
Re-operation	6 months	2,811	1.82
	1 year	3,467	2.24
	2 years	4,428	2.86
	5 years	5,897	3.81
	10 years	7,045	4.56

Table 3-2 SHAR study outcomes

3.3.1.1. Summary statistics of outcomes

Tables 3-3, 3-4, and 3-5 show summary and Kaplan-Meier (KM) estimates for cumulative incidence of mortality, revision, and re-operation outcomes, respectively, at each time point by age and sex. As shown in the tables, the incidence of all outcomes was generally higher in men when compared to women of a similar age.

			30-day death		60-day death		90-day death		1-year death	
		N total	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)
All patients		154,595	164	0.10% (0.06-0.14)	265	0.18% (0.10-0.28)	364	0.24% (0.15-0.34)	1,301	0.98% (0.70-1.25)
Male	All	66,868	92	0.14% (0.09-0.19)	159	0.23% (0.16-0.31)	203	0.29% (0.21-0.37)	680	1.35% (0.94-1.75)
	18-34 years	395	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	10	2.53% (0.97-4.07)
	35-49 years	4,248	0	0.00% (0.00-0.00)	2	0.05% (0.00-0.11)	2	0.05% (0.00-0.11)	8	0.19% (0.06-0.32)
	50-59 years	11,262	4	0.04% (0.00-0.07)	7	0.06% (0.02-0.11)	8	0.07% (0.02-0.12)	34	0.30% (0.20-0.40)
	60-69 years	22,804	14	0.06% (0.03-0.09)	22	0.10% (0.06-0.14)	28	0.12% (0.08-0.17)	124	0.54% (0.45-0.64)
	70-79 years	21,287	33	0.16% (0.10-0.21)	66	0.31% (0.24-0.38)	91	0.43% (0.34-0.52)	283	1.33% (1.18-1.48)
	≥ 80 years	6,872	41	0.60% (0.41-0.78)	62	0.90% (0.68-1.13)	74	1.08% (0.83-1.32)	221	3.22% (2.80-3.63)
Female	All	87,727	72	0.07% (0.03-0.10)	106	0.13% (0.05-0.25)	161	0.19% (0.09-0.32)	621	0.62% (0.47-0.81)
	18-34 years	416	0	0.00% (0.00-0.00)	1	0.24% (0.00-0.71)	1	0.24% (0.00-0.71)	1	0.24% (0.00-0.71)
	35-49 years	3,328	1	0.03% (0.00-0.09)	1	0.03% (0.00-0.09)	3	0.09% (0.00-0.19)	10	0.30% (0.11-0.49)
	50-59 years	10,911	4	0.04% (0.00-0.07)	5	0.05% (0.01-0.09)	8	0.07% (0.02-0.12)	32	0.29% (0.19-0.39)
	60-69 years	27,561	14	0.05% (0.02-0.08)	19	0.07% (0.04-0.10)	27	0.10% (0.06-0.13)	107	0.39% (0.31-0.46)
	70-79 years	32,274	23	0.07% (0.04-0.10)	39	0.12% (0.08-0.16)	61	0.19% (0.14-0.24)	227	0.70% (0.61-0.79)
	≥ 80 years	13,237	30	0.23% (0.15-0.31)	41	0.31% (0.22-0.40)	61	0.46% (0.35-0.58)	244	1.84% (1.61-2.07)

Table 3-3 Summary and Kaplan-Meier (KM) estimates for the cumulative incidence of mortality outcomes at each time point by age and sex using the SHAR

			6-month revision		1-year revision		2-year revision		5-year revision		10-year revision	
		N total	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)
All patients		154,595	1,815	1.18% (0.85-1.51)	2,287	1.54% (1.14-1.93)	3,011	1.98% (1.54-2.42)	4,179	2.97% (2.41-3.54)	5,133	3.89% (3.22-4.55)
Male	All	66,868	1,009	1.45% (1.07-1.82)	1,284	1.90% (1.45-2.35)	1,715	2.43% (1.94-2.91)	2,372	3.65% (3.02-4.28)	2,842	4.61% (3.88-5.33)
	18-34 years	395	5	1.27% (0.16-2.36)	8	2.03% (0.63-3.40)	8	2.03% (0.63-3.40)	16	4.05% (2.09-5.98)	24	6.08% (3.69-8.40)
	35-49 years	4,248	36	0.85% (0.57-1.12)	55	1.29% (0.95-1.63)	84	1.98% (1.56-2.40)	163	3.84% (3.26-4.41)	214	5.04% (4.38-5.69)
	50-59 years	11,262	162	1.44% (1.22-1.66)	215	1.91% (1.66-2.16)	327	2.90% (2.59-3.21)	463	4.11% (3.74-4.48)	560	4.97% (4.57-5.37)
	60-69 years	22,804	323	1.42% (1.26-1.57)	429	1.88% (1.70-2.06)	563	2.47% (2.27-2.67)	806	3.53% (3.29-3.77)	989	4.34% (4.07-4.60)
	70-79 years	21,287	335	1.57% (1.41-1.74)	415	1.95% (1.76-2.14)	551	2.59% (2.37-2.80)	712	3.34% (3.10-3.59)	820	3.85% (3.59-4.11)
	≥ 80 years	6,872	148	2.15% (1.81-2.50)	162	2.36% (2.00-2.72)	182	2.65% (2.27-3.03)	212	3.08% (2.68-3.49)	235	3.42% (2.99-3.85)
Female	All	87,727	806	0.91% (0.64-1.20)	1,003	1.18% (0.83-1.52)	1,296	1.54% (1.15-1.94)	1,807	2.30% (1.80-2.80)	2,291	3.18% (2.57-3.77)
	18-34 years	416	3	0.72% (0.00-1.53)	5	1.20% (0.15-2.24)	7	1.68% (0.44-2.91)	12	2.88% (1.26-4.48)	19	4.57% (2.54-6.55)
	35-49 years	3,328	33	0.99% (0.65-1.33)	41	1.23% (0.86-1.61)	56	1.68% (1.24-2.12)	93	2.79% (2.23-3.35)	137	4.12% (3.44-4.79)
	50-59 years	10,911	101	0.93% (0.75-1.11)	130	1.19% (0.99-1.39)	166	1.52% (1.29-1.75)	245	2.25% (1.97-2.52)	370	3.39% (3.05-3.73)
	60-69 years	27,561	203	0.74% (0.64-0.84)	273	0.99% (0.87-1.11)	376	1.36% (1.23-1.50)	577	2.09% (1.92-2.26)	755	2.74% (2.55-2.93)
	70-79 years	32,274	313	0.97% (0.86-1.08)	386	1.20% (1.08-1.31)	489	1.52% (1.38-1.65)	636	1.97% (1.82-2.12)	752	2.33% (2.17-2.49)
	≥ 80 years	13,237	153	1.16% (0.97-1.34)	168	1.27% (1.08-1.46)	202	1.53% (1.32-1.73)	244	1.84% (1.61-2.07)	258	1.95% (1.71-2.18)

Table 3-4 Summary and Kaplan-Meier (KM) estimates for the cumulative incidence of revision outcomes at each time point by age and sex using the SHAR

			6-month re-operation		1-year re-operation		2-year re-operation		5-year re-operation		10-year re-operation	
		N total	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)
All patients		154,595	2,811	1.92% (1.47-2.36)	3,467	2.48% (1.96-3.00)	4,428	3.07% (2.50-3.65)	5,897	4.35% (3.64-5.04)	7,045	5.43% (4.63-6.21)
Male	All	66,868	1,541	2.43% (1.89-2.94)	1,912	3.19% (2.55-3.82)	2,451	3.85% (3.17-4.52)	3,261	5.33% (4.53-6.11)	3,822	6.46% (5.58-7.33)
	18-34 years	395	14	3.54% (1.70-5.35)	22	5.57% (3.28-7.80)	22	5.57% (3.28-7.80)	30	7.59% (4.95-10.17)	39	9.87% (6.88-12.77)
	35-49 years	4,248	51	1.20% (0.87-1.53)	72	1.69% (1.31-2.08)	108	2.54% (2.07-3.01)	216	5.08% (4.42-5.74)	281	6.61% (5.86-7.36)
	50-59 years	11,262	247	2.19% (1.92-2.30)	317	2.81% (2.51-3.12)	450	4.00% (3.63-4.36)	618	5.49% (5.07-5.91)	733	6.51% (6.05-6.96)
	60-69 years	22,804	481	2.11% (1.92-2.30)	622	2.73% (2.52-2.94)	786	3.45% (3.21-3.68)	1,074	4.71% (4.43-4.98)	1,285	5.63% (5.34-5.93)
	70-79 years	21,287	540	2.54% (2.33-2.75)	655	3.08 % (2.84-3.31)	834	3.92% (3.66-4.18)	1,029	4.83% (4.55-5.12)	1,161	5.45% (5.15-5.76)
	≥ 80 years	6,872	208	3.03% (2.62-3.43)	224	3.26% (2.84-3.68)	251	3.65% (3.21-4.10)	295	4.29% (3.81-4.77)	323	4.70% (4.20-5.20)
Female	All	87,727	1,270	1.42% (1.06-1.78)	1,555	1.78% (1.37-2.19)	1,977	2.30% (1.83-2.78)	2,636	3.38% (2.76-3.98)	3,223	4.40% (3.69-5.10)
	18-34 years	416	5	1.20% (0.15-2.24)	7	1.68% (0.44-2.91)	10	2.40% (0.92-3.86)	19	4.57% (2.54-6.55)	27	6.49% (4.09-8.83)
	35-49 years	3,328	46	1.38% (0.98-1.78)	60	1.80% (1.35-2.25)	81	2.43% (1.91-2.96)	126	3.79% (3.14-4.43)	175	5.26% (4.50-6.01)
	50-59 years	10,911	139	1.27% (1.06-1.48)	178	1.63% (1.39-1.87)	222	2.03% (1.77-2.30)	324	2.97% (2.65-3.29)	469	4.30% (3.92-4.68)
	60-69 years	27,561	317	1.15% (1.02-1.28)	405	1.47% (1.33-1.61)	561	2.04% (1.87-2.20)	814	2.95% (2.75-3.15)	1,022	3.71% (3.48-3.93)
	70-79 years	32,274	499	1.55% (1.41-1.68)	613	1.90% (1.75-2.05)	762	2.36% (2.20-2.53)	948	2.94% (2.75-3.12)	1,094	3.39% (3.19-3.59)
	≥ 80 years	13,237	264	1.99% (1.76-2.23)	292	2.21% (1.96-2.46)	341	2.58% (2.31-2.85)	405	3.06% (2.77-3.35)	436	3.29% (2.99-3.60)

Table 3-5 Summary and Kaplan-Meier (KM) estimates for the cumulative incidence of re-operation outcomes at each time point by age and sex using the SHAR

Despite the overall increase in the number of hip arthroplasty procedures performed in Sweden during the studied period, there was a gradual reduction in the risk of death over the ten-year period. Notably, there was a statistically significant decrease in mortality rates from year 2013 onwards when compared to the reference group of year 2008 (KM estimate for one-year mortality 1.17% in 2008, 0.77% - 0.89% between 2013 and 2017) and after

adjustment for age and sex, Hazard Ratio (HR) decreased by over one-third (for 2017 vs 2000-2008 HR 0.62%, 95% CI 0.48-0.79; p<0.001; Figure 3-1; Table 3-6).

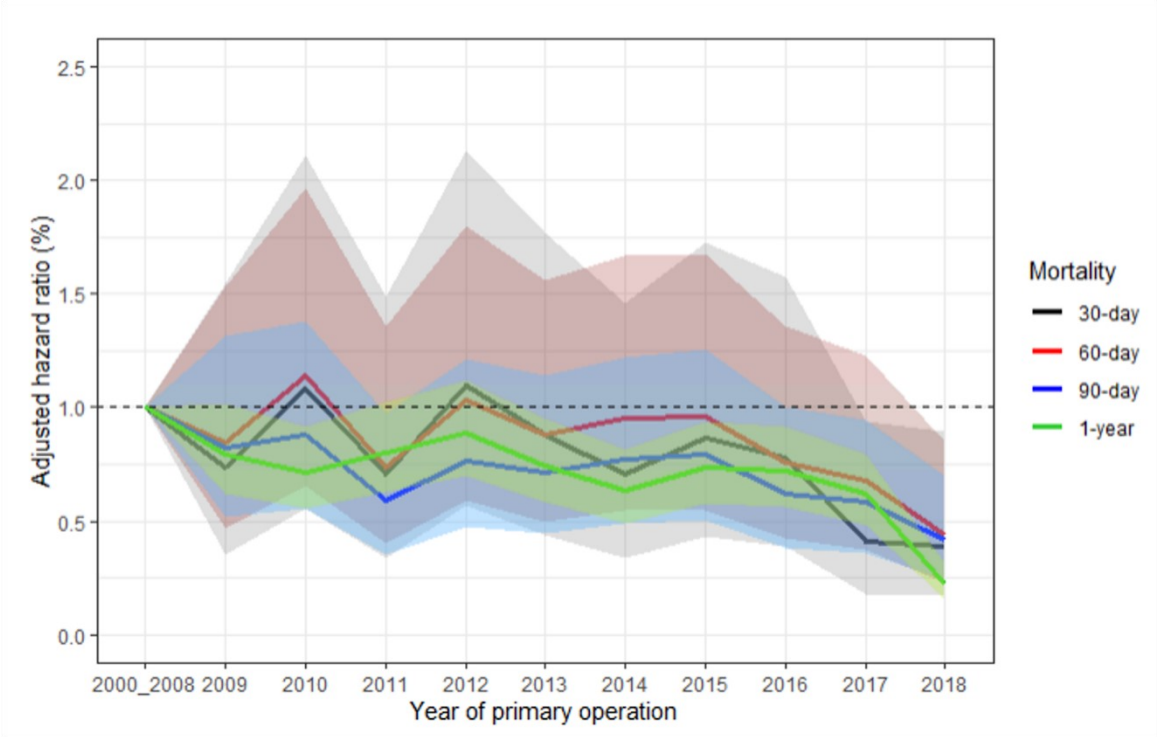


Figure 3-1 Change in mortality rates after primary hip arthroplasty in the SHAR between 2008 and 2018

		30-day death			60-day death			90-day death			1-year death		
Year	N total	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§
2000-2008	11,155	15	0.13%	1.00	22	0.20%	1.00	36	0.32%	1.00	131	1.17%	1.00
2009	13,027	13	0.10%	0.73 (0.35-1.54)	22	0.17%	0.84 (0.46-1.52)	35	0.27%	0.82 (0.51-1.31)	122	0.94%	0.79 (0.61-1.01)
2010	13,546	20	0.15%	1.07 (0.55-2.10)	31	0.23%	1.13 (0.65-1.96)	39	0.29%	0.87 (0.55-1.37)	115	0.85%	0.71 (0.55-0.91)**
2011	13,697	13	0.09%	0.70 (0.33-1.48)	20	0.15%	0.73 (0.40-1.35)	26	0.19%	0.58 (0.35-0.97)*	129	0.94%	0.80 (0.63-1.02)
2012	13,946	21	0.15%	1.10 (0.56-2.13)	29	0.21%	1.03 (0.59-1.79)	35	0.25%	0.76 (0.47-1.21)	148	1.06%	0.88 (0.70-1.12)
2013	13,989	17	0.12%	0.88 (0.44-1.76)	25	0.18%	0.88 (0.49-1.56)	33	0.24%	0.71 (0.44-1.14)	125	0.89%	0.74 (0.58-0.95)*
2014	14,226	14	0.10%	0.70 (0.33-1.45)	28	0.20%	0.95 (0.54-1.66)	37	0.26%	0.72 (0.48-1.22)	110	0.77%	0.63 (0.49-0.81)***
2015	14,403	17	0.12%	0.86 (0.43-1.72)	28	0.19%	0.96 (0.54-1.67)	38	0.26%	0.79 (0.50-1.25)	127	0.88%	0.73 (0.57-0.93)*
2016	14,995	16	0.11%	0.77 (0.38-1.57)	23	0.15%	0.75 (0.42-1.35)	31	0.21%	0.61 (0.38-1.00)	130	0.87%	0.71 (0.56-0.91)**
2017	15,512	9	0.06%	0.41 (0.17-0.93)*	22	0.14%	0.67 (0.37-1.22)	31	0.20%	0.58 (0.36-0.94)*	119	0.77%	0.62 (0.48-0.79)***
2018	16,099	9	0.06%	0.39 (0.17-0.89)*	15	0.09%	0.44 (0.22-0.85)*	23	0.14%	0.41 (0.24-0.70)***	45	0.28%	0.22 (0.16-0.31)***

Table 3-6 Change in mortality rates by year of primary operation in the SHAR
§ Adjusted for Age and Gender; * p<0.05, ** p<0.01, *** p<0.001
KM = Kaplan-Meier; HR = Hazard Ratio; CI = Confidence Interval

Changes in revision and re-operation rates by year of primary hip arthroplasty procedure are highlighted in Figures 3-2 and 3-3 (tables are presented in Appendix 3, A3.1 and A3.2), respectively. There was a steady increase in hazard ratio risk for revision between the years 2010 and 2015 with a significant increase in year 2016 (Figure 3-2), whilst these changes were less pronounced for re-operations (Figure 3-3).

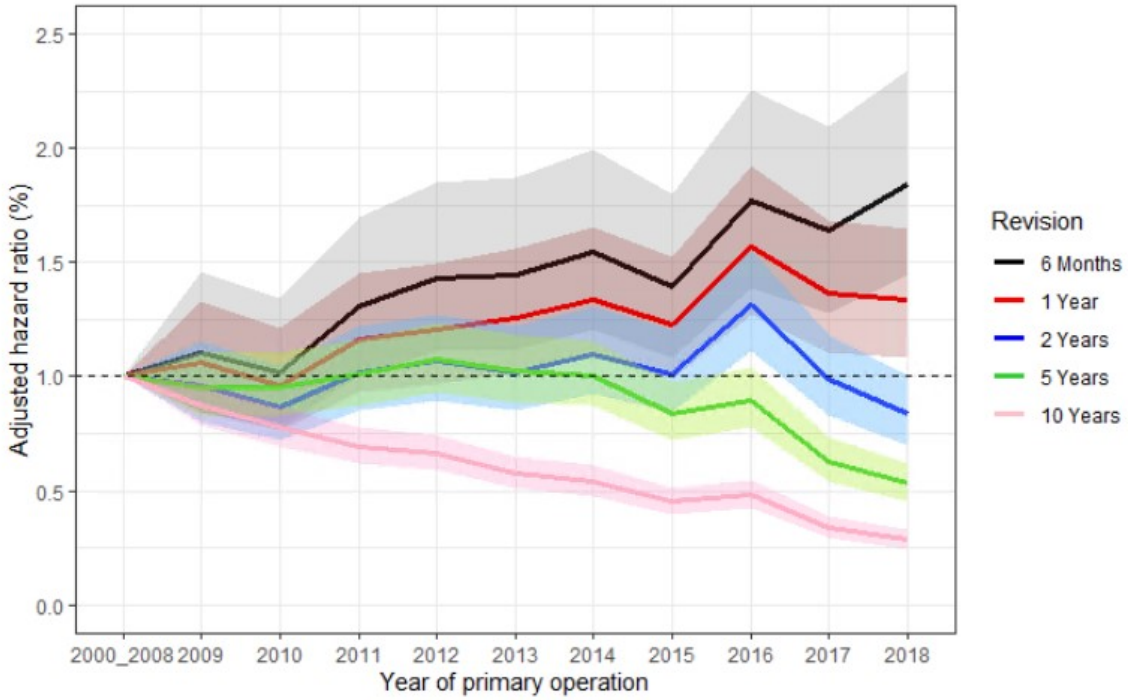


Figure 3-2 Change in revision rates after primary hip arthroplasty in the SHAR between 2008 and 2018

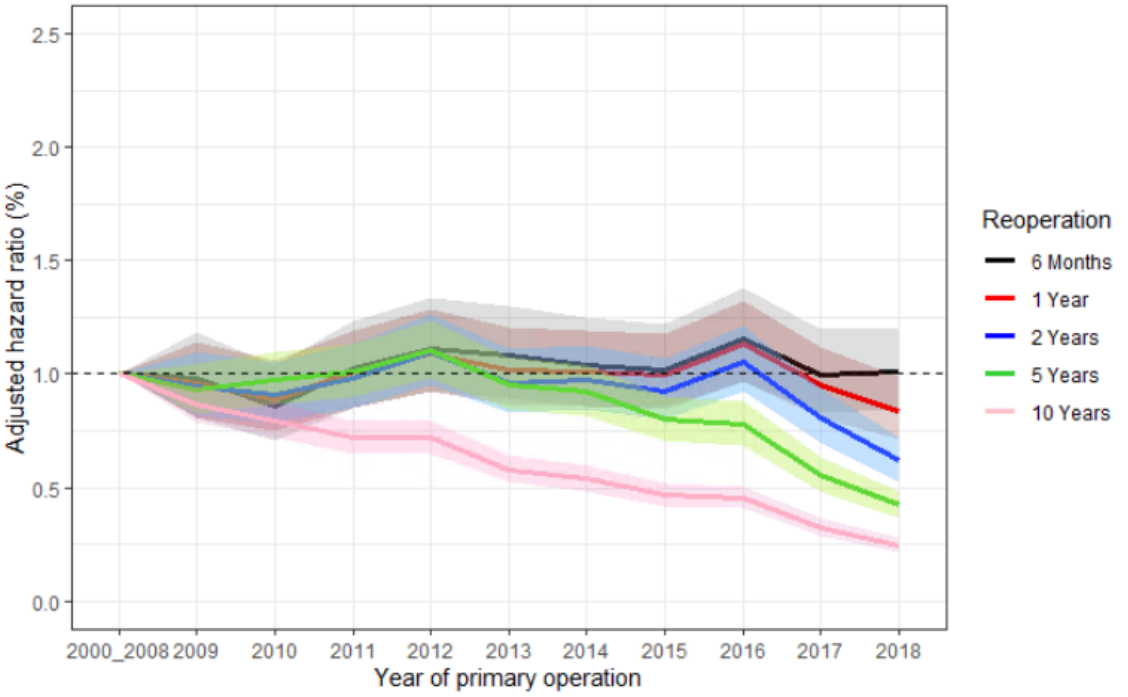


Figure 3-3 Change in re-operation rates after primary operation in the SHAR between 2008 and 2018

3.3.1.2. Machine Learning models training and internal validation

The best-performing ML algorithms in the internal validation cohort for each time-defined outcome were selected as the SHAR SafeHip models (Table 3-7). Prediction performance of all applied machine learning algorithms in the internal validation cohort are presented in Table 3-8. In the prediction of mortality outcomes, the Gradient Boosting Machine (GBM), Random Forest (RF), Ridge regression, and Lasso regression consistently ranked among the top-performing models. These models exhibited a remarkably similar area under the curve (AUC) of approximately 0.80. However, when it came to predicting 30-day and 60-day mortalities, Ridge regression outperformed the other models by a narrow margin. On the other hand, for predicting 90-day and 1-year mortalities, GBM demonstrated the best performance.

When it came to predicting implant failure outcomes, the Random Forest (RF) algorithm displayed superior performance compared to the other competing algorithms. It achieved area under the curve (AUC) values of 0.70 or higher in predicting all re-operation and revision outcomes. Importantly, all the algorithms exhibited comparable performances in the training cohort indicating good generalisability and minimal overfitting.

SHAR SafeHip Scores		
Outcome	Model name	AUC [CI]
60-day Mortality	Ridge Regression	0.81 [0.75 – 0.87]
1-year Re-operation 2-year Re-operation	Random Forest	0.77 [0.75 – 0.79]
1-year Revision	Random Forest	0.72 [0.70 – 0.74]

Table 3-7 SHAR SafeHip Scores
the best-performing machine learning algorithms in the internal validation cohort for the selected time-defined outcome

Outcome	Model performance: AUC in % (95% CI)						
	Random forest	Gradient boosting machine	Ridge regression	Lasso regression	Logistic regression	Classification tree	Classification tree with pruning
Mortality							
30 days	0.73 (0.65-0.81)	0.77 (0.70-0.85)	**0.79 (0.71-0.86)**	0.77 (0.69-0.85)	0.77 (0.68-0.85)	0.72 (0.64-0.80)	0.72 (0.64-0.80)
60 days	0.78 (0.71-0.84)	0.81 (0.75-0.86)	**0.81 (0.75-0.86)**	0.80 (0.75-0.87)	0.79 (0.72-0.86)	0.75 (0.69-0.82)	0.75 (0.69-0.82)
90 days	0.78 (0.72-0.85)	**0.80 (0.74-0.86)**	0.80 (0.74-0.85)	0.80 (0.73-0.85)	0.79 (0.72-0.85)	0.75 (0.69-0.81)	0.75 (0.69-0.81)
1 year	0.77 (0.74-0.80)	**0.77 (0.75-0.80)**	0.76 (0.74-0.80)	0.77 (0.74-0.80)	0.77 (0.74-0.80)	0.71 (0.69-0.74)	0.71 (0.69-0.74)
Revision							
6 months	**0.71 (0.70-0.71)**	0.68 (0.66-0.71)	0.68 (0.65-0.70)	0.68 (0.65-0.70)	0.68 (0.65-0.70)	0.56 (0.54-0.58)	0.56 (0.54-0.58)
1 year	**0.72 (0.69-0.74)**	0.68 (0.65-0.70)	0.67 (0.64-0.69)	0.67 (0.64-0.69)	0.67 (0.64-0.69)	0.64 (0.61-0.66)	0.63 (0.61-0.65)
2 years	**0.72 (0.70-0.74)**	0.67 (0.65-0.69)	0.65 (0.63-0.68)	0.65 (0.63-0.68)	0.66 (0.63-0.68)	0.63 (0.61-0.65)	0.61 (0.59-0.64)
5 years	**0.71 (0.69-0.73)**	0.66 (0.64-0.68)	0.65 (0.63-0.67)	0.65 (0.63-0.67)	0.65 (0.63-0.67)	0.62 (0.61-0.64)	0.60 (0.59-0.62)
10 years	**0.72 (0.70-0.74)**	0.66 (0.64-0.68)	0.65 (0.63-0.67)	0.65 (0.63-0.67)	0.65 (0.63-0.67)	0.62 (0.61-0.64)	0.59 (0.58-0.61)
Re-operation							
6 months	**0.73 (0.71-0.76)**	0.69 (0.66-0.71)	0.68 (0.65-0.70)	0.68 (0.66-0.70)	0.68 (0.66-0.70)	0.56 (0.55-0.58)	0.56 (0.55-0.58)
1 year	**0.77 (0.75-0.79)**	0.68 (0.66-0.70)	0.67 (0.65-0.69)	0.67 (0.65-0.69)	0.67 (0.65-0.69)	0.64 (0.62-0.66)	0.63 (0.61-0.65)
2 years	**0.77 (0.75-0.79)**	0.68 (0.66-0.70)	0.66 (0.64-0.68)	0.66 (0.64-0.68)	0.66 (0.64-0.68)	0.63 (0.61-0.65)	0.62 (0.60-0.63)
5 years	**0.75 (0.73-0.77)**	0.67 (0.65-0.68)	0.65 (0.64-0.67)	0.65 (0.64-0.67)	0.65 (0.64-0.67)	0.62 (0.60-0.63)	0.59 (0.58-0.61)
10 years	**0.75 (0.74-0.77)**	0.67 (0.66-0.68)	0.66 (0.64-0.67)	0.66 (0.64-0.67)	0.66 (0.64-0.67)	0.63 (0.62-0.64)	0.59 (0.58-0.60)

Table 3-8 Prediction performance of all applied machine learning algorithms in the internal validation cohort. *The best-performing models for each outcome are highlighted in bold font*
AUC = area under the receiver operating characteristic curve, CI = confidence interval

3.3.2. National Joint Registry

Within 30 days and 90 days following the primary operation, the number of deaths was 0.11% and 0.24, respectively, figures comparable to the SHAR derivation dataset (Table 3-9). Whilst re-operation did not form an outcome from the NJR dataset, the revision rates

in the NJR dataset were considerably lower and approximately half that of the SHAR counterpart.

Outcome	Time from primary operation	NJR Derivation cohort (N=355,933) N	%	SHAR Derivation cohort (N=154,595) %
Mortality	30 days	506	0.14	0.11
	60 days	782	0.21	0.17
	90 days	1,047	0.29	0.24
	1 year	3,636	1.02	0.84
Revision	6 months	1,824	0.51	1.17
	1 year	2,450	0.68	1.48
	2 years	3,362	0.94	1.95
	5 years	4,671	1.31	2.70
	10 years	5,195	1.45	3.32

Table 3-9 NJR study outcomes

3.3.2.1. Summary statistics of outcomes

Tables 3-10 and 3-11 show summary and Kaplan-Meier (KM) estimates for cumulative incidence of mortality and revision outcomes, respectively, at each time point by age and sex. Similar to the SHAR findings and as shown in the tables, the incidence of all outcomes was generally higher in men when compared to women of a similar age.

			30-day death		60-day death		90-day death		1-year death	
		N total	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)
All patients		355,933	506	0.12% (0.09-0.16)	782	0.19% (0.15-0.24)	1047	0.27% (0.21-0.41)	3636	0.96% (0.84-1.23)
Male	All	141,677	276	0.16% (0.13-0.22)	412	0.25% (0.20-0.31)	538	0.36% (0.28-0.59)	1786	1.16% (1.03-1.45)
	18-34 years	632	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	1	0.16% (0.02-1.15)	1	0.16% (0.02-1.15)
	35-49 years	8,031	1	0.01% (0.00-0.09)	3	0.04% (0.01-0.12)	7	0.09% (0.04-0.19)	27	0.37% (0.25-0.53)
	50-59 years	21,124	9	0.04% (0.02-0.08)	14	0.07% (0.04-0.11)	21	0.10% (0.07-0.16)	94	0.49% (0.40-0.59)
	60-69 years	43,932	32	0.07% (0.05-0.10)	61	0.14% (0.11-0.18)	84	0.19% (0.16-0.24)	305	0.75% (0.67-0.84)
	70-79 years	47,997	97	0.20% (0.17-0.25)	142	0.30% (0.25-0.35)	180	0.38% (0.33-0.44)	671	1.51% (1.40 -1.63)
	≥ 80 years	19,961	137	0.69% (0.58-0.82)	192	0.97% (0.84-1.12)	245	1.25% (1.10-1.41)	688	3.72% (3.46-4.01)
Female	All	214,256	230	0.08% (0.06-0.11)	370	0.13 % (0.10-0.17)	509	0.18% (0.15-0.23)	1850	0.76% (0.66-1.01)
	18-34 years	651	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	1	0.17% (0.02-1.19)
	35-49 years	8,837	3	0.03% (0.01-0.11)	4	0.05% (0.02-0.12)	7	0.08% (0.04-0.17)	31	0.38% (0.27-0.54)
	50-59 years	25,996	5	0.02% (0.01-0.05)	10	0.04% (0.02-0.07)	16	0.06% (0.04-0.10)	73	0.31% (0.24-0.38)
	60-69 years	60,383	20	0.03% (0.02-0.05)	34	0.06% (0.04-0.08)	55	0.09% (0.07-0.12)	244	0.44% (0.38-0.49)
	70-79 years	79,030	81	0.10% (0.08-0.13)	137	0.18% (0.15-0.21)	180	0.23% (0.20-0.27)	612	0.84% (0.77-0.91)
	≥ 80 years	39,359	121	0.31% (0.26-0.37)	185	0.48% (0.41-0.55)	251	0.65% (0.57-0.74)	889	2.44% (2.28-2.60)

Table 3-10 Summary and Kaplan-Meier (KM) estimates for the cumulative incidence of mortality outcomes at each time point by age and sex using the NJR

			6-month revision		1-year revision		2-year revision		5-year revision		10-year revision	
		N total	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)
All patients		355,933	1824	0.53% (0.41-0.79)	2,450	0.74% (0.60-1.03)	3,362	1.11% (0.92-1.44)	4,671	2.05% (1.69-2.62)	5,195	3.45% (2.63-4.66)
Male	All	141,677	836	0.64% (0.48-0.92)	1,145	0.90% (0.72-1.24)	1,558	1.34% (1.08-1.74)	2,134	2.30% (1.90-2.88)	2,373	4.01% (3.01-5.48)
	18-34 years	632	5	0.80% (0.33-1.92)	7	1.14% (0.55-2.38)	10	1.74% (0.93-3.21)	13	2.72% (1.53-4.81)	16	5.00% (2.72-9.10)
	35-49 years	8,031	44	0.56% (0.42-0.75)	68	0.89% (0.71-1.13)	101	1.43% (1.17-1.73)	168	3.08% (2.63-3.60)	181	4.12% (3.38-5.00)
	50-59 years	21,124	115	0.56% (0.46-0.67)	176	0.88% (0.76-1.02)	252	1.34% (1.19-1.52)	352	2.28% (2.04-2.54)	404	5.25% (3.76-7.31)
	60-69 years	43,932	228	0.53% (0.47-0.60)	320	0.76% (0.69-0.85)	451	1.14% (1.04-1.25)	614	1.86% (1.71-2.02)	683	2.96% (2.61-3.36)
	70-79 years	47,997	294	0.63% (0.56-0.71)	390	0.86% (0.78-0.95)	522	1.21% (1.11-1.32)	700	1.95% (1.81-2.11)	780	3.63% (3.04-4.33)
	≥ 80 years	19,961	150	0.78% (0.66-0.91)	184	0.97% (0.84-1.12)	222	1.23% (1.08-1.41)	287	1.95% (1.72-2.21)	309	3.13% (2.58-3.81)
Female	All	214,256	988	0.43% (0.34-0.66)	1,305	0.59% (0.49-0.83)	1,804	0.88% (0.76-1.15)	2,537	1.81% (1.48-2.36)	2,822	2.90% (2.25-3.85)
	18-34 years	651	1	0.15% (0.02-1.09)	1	0.15% (0.02-1.09)	1	0.15% (0.02-1.09)	8	2.11% (1.02-4.35)	8	2.11% (1.02-4.35)
	35-49 years	8,837	49	0.57% (0.43-0.75)	75	0.90% (0.72-1.12)	114	1.45% (1.21-1.74)	165	2.54% (2.17-2.98)	189	4.55% (3.61-5.72)
	50-59 years	25,996	106	0.42% (0.34-0.50)	152	0.61% (0.52-0.72)	230	1.00% (0.88-1.13)	344	1.82% (1.63-2.04)	377	3.23% (2.39-4.36)
	60-69 years	60,383	254	0.43% (0.38-0.49)	341	0.59% (0.53-0.66)	506	0.93% (0.86-1.02)	706	1.56% (1.45-1.69)	810	2.97% (2.60-3.40)
	70-79 years	79,030	377	0.49% (0.44-0.54)	493	0.65% (0.60-0.71)	652	0.91% (0.84-0.98)	933	1.61% (1.50-1.72)	1,026	2.61% (2.26-3.00)
	≥ 80 years	39,359	201	0.52% (0.46-0.60)	243	0.64% (0.57-0.73)	301	0.84% (0.75-0.94)	381	1.25% (1.13-1.40)	412	1.94% (1.63-2.31)

Table 3-11 Summary and Kaplan-Meier (KM) estimates for the cumulative incidence of revision outcomes at each time point by age and sex using the NJR

Like the SHAR findings, there was an overall increase in the number of hip replacements performed in England and Wales per year, coupled with a gradual decrease in the Kaplan-Meier estimate of death at 30 and 60 days. KM estimates at 90 days and 1 year did not seem to change between the years 2009 and 2017 (Table3-12). Hazard ratios and p values were not calculated for the NJR dataset for pragmatic reasons.

		30-day death		60-day death		90-day death		1-year death	
Year	N total	N	KM estimate	N	KM estimate	N	KM estimate	N	KM estimate
2009	17,007	30	0.18%	52	0.31%	66	0.39%	236	1.39%
2010	21,228	35	0.17%	54	0.25%	75	0.35%	262	1.24%
2011	25,676	41	0.16%	60	0.23%	76	0.30%	277	1.08%
2012	32,802	48	0.15%	77	0.24%	104	0.32%	368	1.13%
2013	38,137	54	0.14%	87	0.23%	121	0.32%	440	1.16%
2014	44,587	59	0.13%	96	0.22%	126	0.28%	494	1.11%
2015	45,850	76	0.17%	108	0.24%	147	0.32%	502	1.10%
2016	48,091	64	0.13%	101	0.21%	143	0.30%	494	1.03%
2017	50,180	65	0.13%	98	0.20%	128	0.26%	475	1.01%
2018	32,375	34	0.11%	49	0.17%	61	0.22%	88	0.51%

Table 3-12 Change in mortality rates by year of primary operation in the NJR
KM = Kaplan-Meier

Changes in revision rates by year of primary hip arthroplasty procedure are highlighted in Table 3-13. Short-term revision rates appeared to be constant among the years with a decline in years 2017 and 2018.

		6-month revision		1-year revision		2-year revision		5-year revision		10-year revision	
Year	N total	N	KM estimate	N	KM estimate	N	KM estimate	N	KM estimate	N	KM estimate
2009	17,007	69	0.41%	97	0.57%	167	0.99%	345	2.11%	535	3.62%
2010	21,228	94	0.44%	156	0.74%	254	1.21%	412	2.00%	538	2.85%
2011	25,676	109	0.43%	157	0.61%	271	1.07%	458	1.84%	580	2.48%
2012	32,802	149	0.46%	210	0.64%	320	0.98%	553	1.73%	621	2.02%
2013	38,137	167	0.44%	261	0.69%	375	0.99%	604	1.64%	622	1.77%
2014	44,587	235	0.53%	334	0.75%	457	1.03%	640	1.51%	640	1.51%
2015	45,850	295	0.64%	366	0.80%	488	1.07%	606	1.55%	606	1.55%
2016	48,091	274	0.57%	374	0.78%	512	1.09%	535	1.22%	535	1.22%
2017	50,180	291	0.58%	351	0.72%	374	0.81%	374	0.81%	374	0.81%
2018	32,375	141	0.56%	144	0.63%	144	0.63%	144	0.63%	144	0.63%

Table 3-13 Change in revision rates by year of primary operation in the NJR
KM = Kaplan-Meier

3.3.2.2. Machine Learning models training and external validation

The best-performing ML algorithms in the NJR external validation cohort were selected as the final NJR SafeHip models (Table 3-14). Prediction performance of all applied machine learning algorithms in the external validation cohort are presented in Table 3-15. In the prediction of mortality outcomes, Lasso regression and the Gradient Boosting Machine (GBM) consistently ranked among the top-performing models. Similar to the SHAR SafeHip Score, the NJR models exhibited a remarkably similar area under the curve (AUC) of approximately 0.80. Lasso regression was the best performing model in predicting 30-day and 60-day mortalities, whilst GBM demonstrated the best performance in predicting 90-day and 1-year mortalities.

NJR SafeHip Scores		
Outcome	Model name	AUC [CI]
30-day Mortality	Lasso Regression	0.82 [0.77 – 0.86]
6-month Revision	Ridge Regression	0.61 [0.58 – 0.64]

Table 3-14 NJR SafeHip Scores
the best-performing machine learning algorithms in the external validation cohort for the selected time-defined outcome

When it came to predicting implant revision outcomes, Ridge regression and the Random Forest (RF) algorithm displayed superior performance compared to other machine learning models. However, it achieved lower AUCs in comparison to the SHAR revision predictive models with values of around 0.60 for all time points. Importantly, all the algorithms exhibited comparable performances in the training cohort indicating good generalisability and minimal overfitting.

Outcome	Model performance: AUC in % (95% CI)					
	Random forest	Gradient boosting machine	Ridge regression	Lasso regression	Logistic regression	Classification tree
Mortality						
30 days	0.817 (0.77-0.85)	0.819 (0.77-0.86)	0.817 (0.77-0.86)	**0.820 (0.77-0.86)**	0.818 (0.77-0.86)	0.754 (0.70-0.80)
60 days	0.791 (0.75-0.82)	0.798 (0.76-0.83)	0.797 (0.76-0.83)	**0.803 (0.76-0.83)**	0.800 (0.76-0.83)	0.741 (0.70-0.77)
90 days	0.806 (0.77-0.82)	**0.807 (0.77-0.83)**	0.798 (0.76-0.82)	0.800 (0.77-0.83)	0.800 (0.77-0.82)	0.738 (0.70-0.76)
1 year	0.760 (0.74-0.78)	**0.770 (0.75-0.79)**	0.760 (0.74-0.78)	0.760 (0.74-0.78)	0.760 (0.74-0.78)	0.720 (0.70-0.74)
Revision						
6 months	0.605 (0.57-0.63)	0.601 (0.57-0.63)	**0.611 (0.58-0.64)**	0.608 (0.57-0.63)	0.600 (0.57-0.63)	0.560 (0.53-0.58)
1 year	**0.599 (0.57-0.62)**	0.592 (0.56-0.61)	0.596 (0.57-0.62)	0.594 (0.56-0.61)	0.580 (0.56-0.61)	0.570 (0.54-0.59)
2 years	**0.603 (0.57-0.62)**	0.590 (0.57-0.62)	0.602 (0.57-0.62)	0.598 (0.57-0.62)	0.598 (0.57-0.62)	0.563 (0.55-0.58)
5 years	0.599 (0.57-0.62)	0.600 (0.57-0.62)	**0.606 (0.57-0.63)**	0.605 (0.57-0.63)	0.600 (0.57-0.63)	0.550 (0.52-0.57)

Table 3-15 Prediction performance of all applied machine learning algorithms in the external validation cohort. *The best-performing models for each outcome are highlighted in bold font*
AUC = area under the receiver operating characteristic curve, CI = confidence interval

3.3.3. Bland-Altman Method

The Bland-Altman technique was employed to evaluate the concordance between the AUCs produced by the NJR and SHAR machine learning models. The Bland–Altman plot for mortality (Figure 3-4) showed a bias of 0.012 AUCs (95% CI 0.000 to 0.022). The upper and lower limits of agreement were 0.062 and -0.039 AUCs, respectively.

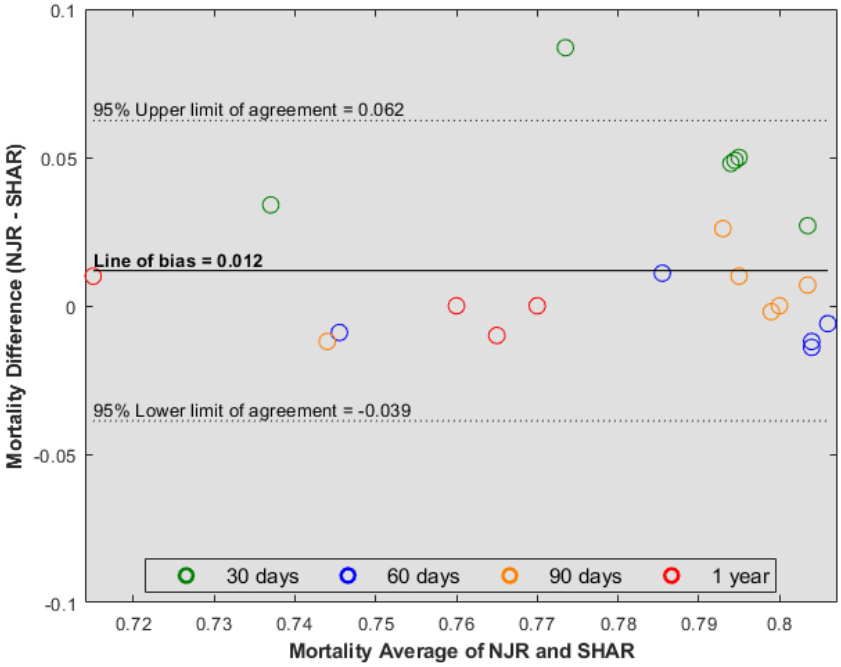


Figure 3-4 Limits of agreement for the areas under the curve (AUCs) between NJR and SHAR mortality machine learning models

In terms of revision, the Bland-Altman plot (Figure 3-5) showed a bias of -0.072 AUCs (95% CI -0.083 to -0.060). The upper and lower limits of agreement were -0.020 and -0.125 AUCs, respectively.

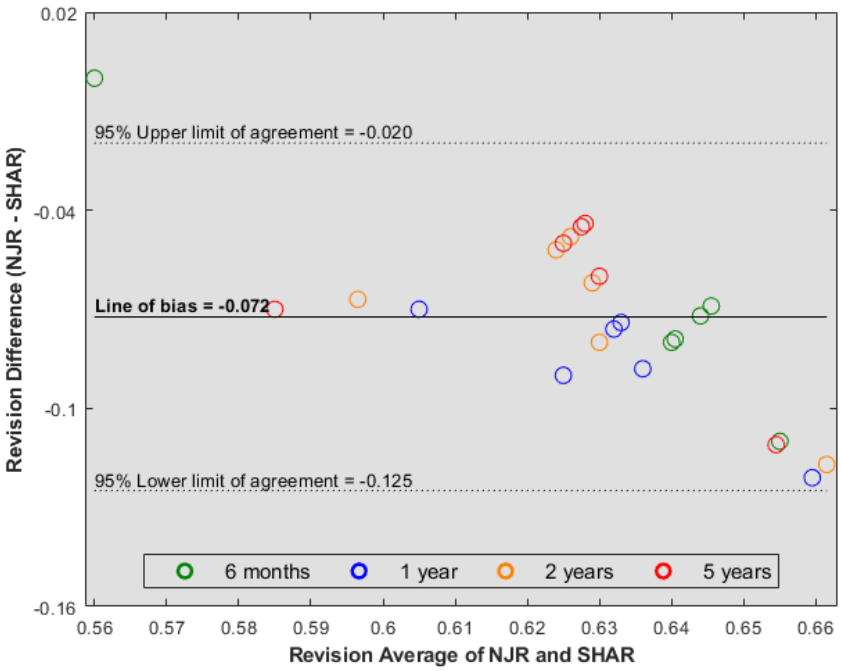


Figure 3-5 Limits of agreement for the areas under the curve (AUCs) between NJR and SHAR revision machine learning models

3.4. Discussion

3.4.1. Summary of results

This study has demonstrated that machine learning based predictive models for mortality, revision, and re-operation after primary hip arthroplasty achieve a slightly higher level of accuracy, as determined by the area under the curve (AUC), compared to conventional logistic regression models.

Descriptive statistics of the SHAR and NJR datasets revealed temporal and gender variations related to the incidences of mortality, revision, and re-operation. Much like the SHAR and the NJR annual reports, younger patients had a lower risk of death but a higher risk of long-term "implant failure", or revision surgery (5, 6). For mortality, the differences were primarily apparent in the first 90-days postoperatively and persisted to one-year. Across the years, the NJR showed a steady decline in revision rates whilst the SHAR data exhibited fluctuations with a significant decline in 2015; this is likely attributed to the reduction in the use of metal-on-metal hip arthroplasty following regulatory concerns related to metal debris and ions in Sweden (373). For similar age group comparisons, the risk of death was approximately twice as high in males at one-year (SHAR) and ninety days (NJR), however, the proportion of males that are revised or re-operated on is generally higher than females with the variations growing more pronounced with the passage of time. As discussed in Section 1.9.1, there is established evidence of no significant association between gender and mortality or revision outcomes following primary total hip arthroplasty. Therefore, these findings highlight the need for statistical adjustment for age and gender, as addressed in Chapter 4.

The best performing predictive models, as illustrated by the SafeHip scores, demonstrated good predictive abilities (AUCs of 0.80 or higher) for the prediction of death risks using a

set of patient-, surgical-, and implant-specific variables. In both datasets (SHAR and NJR), Gradient Boosting Machine was the best performing ML model in predicting 90-day and 1-year mortality, whilst Ridge regression (SHAR) and Lasso regression (NJR) algorithms outperformed other models in predicting 60-day and 30-day death, respectively. Predicting revision and re-operation risks using machine learning algorithms was more challenging with AUCs of above 0.70 in the SHAR dataset and around 0.60 in the NJR dataset. The Random Forest (RF) model was the best performing algorithm in the SHAR dataset to forecast both revision and re-operation risks. On the other hand, RF and Ridge regression outperformed other models in the NJR dataset. This is in agreement with previous literature highlighting the superior performance of machine-learning algorithms compared to generalised linear statistical models (374-376).

3.4.1.1. Evidence related to mortality predictive models

A previous study developed a Lasso predictive model for 90-day mortality following primary THA using the SHAR dataset for training and the NJR database for external validation (377). The authors reported AUC = 0.75 (95% CI 0.73 to 0.76). However, the authors restricted their training model to a smaller cohort (n_{SHAR} = 53,099, n_{NJR} = 125,428) operated with all-cemented fixations, which significantly impact validity to other fixation types. Similarly, Wilkinson and colleagues used flexible parametric survival regression to provide an individualised estimation of one-year mortality risk after elective hip and knee arthroplasty using the NJR as a training platform and externally validated their algorithms using the Norwegian Arthroplasty Register (NAR) (378). For the hip, they reported a Harrell's concordance index of approximately 0.76, indicating good discriminative ability. Nonetheless, their methodology had several limitations primarily the absence of patient censoring in survivorship modelling, and the external validation data incorporated fewer training variables, which both affect the precision and generalisability of predictions. A

team in the United States used an artificial neural network approach on observations from the national quality improvement programme (n = 77,145) to identify the top five determinants of 30-day mortality following primary THR and reported an AUC of 0.80% (379). Again, the dataset used assessed complications for up to one-month postoperatively, which likely led to under-representing the studied outcomes, in addition to the data being of limited generalisability.

3.4.1.2. Evidence related to revision prediction

Two studies developed predictive models to forecast the risks of revision and re-revision following primary total hip arthroplasty (380, 381). Klemt *et al.* were able to achieve a discriminative ability of AUCs of 0.82 – 0.87 for all-case revision (380) and 0.80 – 0.86 for re-revision (381) following THR. However, their models were developed on data derived from a single tertiary referral centre in the United States; had a limited sample size ($n_{revision}$ = 566, $n_{re\text{-}revision}$ = 408); and did not include disease severity, pre-operative diagnosis, or implant-specific factors in their models – all of which were observed to be leading players in driving our predictions.

Our machine learning models appear to be effective to predict personalised risks of 30-day and 60-day mortality and 6-month and 1-year revision risks. Compared to the AUCs measured by the machine learning models using the NJR dataset, the counterpart in the SHAR dataset showed good agreement for both mortality and revision. The results imply that the SHAR models tend to slightly underestimate the predictions for mortality, with an observed bias of 0.012 AUCs; and slightly overestimate the predictions for revision, with an observed bias of -0.072 AUCs. Furthermore, the even spread of data points across the Bland–Altman plot suggests no systematic difference across the AUC measurements. The limits of agreement 0.062 and -0.039 AUCs for mortality; and -0.020 and -0.125 AUCs for revision were also relatively narrow, and indicate that for approximately 95% of cases,

the predictions measured by the NJR will minimally differ from the predictions generated by the SHAR. This strengthens our confidence in the generalisability of the developed machine learning models. The usefulness of these to support and personalise the shared-decision process on the choice of primary hip arthroplasty should be regarded as a feasibility study at this stage. Randomised trials are needed to examine the validity and practicality of implementing these scores in clinical practice.

3.4.2. Strengths and limitations of this study

The strengths of this study include the large study population, using two of the largest and most detailed hip arthroplasty registries in the world, which enabled adjustment for potential confounding factors. Employing an unselected sample derived from registry datasets enhances the study's external validity. Using different linear and non-linear machine learning techniques in forecasting various outcomes overcame the assumption of linearity that is inherent in traditional statistical methods. The baseline descriptive statistics for outcomes in both the NJR and the SHAR were comparable, indicating similar characteristics and distributions of the outcomes. These comparable statistics limit potential biases that may arise from differences between the two datasets.

Limitations include the observational retrospective nature and the associated inherent biases of registry data, such as selection bias and missing values, as well as the limitations concerning the use of revision/re-operation as endpoint outcomes. However, our study showed good discrimination in the 20% validation cohort and investigated outcomes of four to five different time points. Similarly, our models were compared using the area under the curve (AUC) metric, which is regarded as the gold-standard in evaluating classifier performance across a range of thresholds and allows for comparisons of multiple machine learning models. Nonetheless, the AUC-ROC curve-based metric neglects to address the balance of true positives and true negatives, a nuanced aspect that alternative

measures, like Matthews Correlation Coefficient, precision, and recall, seek to address. Albeit the evaluation of such measures is contingent upon setting a range of thresholds, a practice known to yield variation in clinical practice. Finally, advanced machine learning techniques such as deep learning, neural networks and support vector machines, which are particularly effective in handling complex non-linear relationships were not used in this study since they introduce challenges in terms of model interpretability and explainability. Chapter 4 attempts to identify the key predictors associated with our safety outcomes using CoxNet, a survival analysis machine learning method. Limitations of this chapter in the context of the work as a whole are discussed further in Chapter 10.

3.4.3. Conclusions

This study represents the largest, registry-based machine learning prediction of safety and failure outcomes in primary hip arthroplasty. The findings of superior predictive ability associated with machine learning techniques suggest that ML can provide valuable insights and enhance the decision-making in hip arthroplasty and may uncover non-linear and high-dimensional associations that may be missed by conventional statistical methods. However, researchers should consider the trade-off between model performance and interpretability for a given study.

The next three chapters of this thesis concern the factors affecting the safety, failure rate, and functional outcomes of hip arthroplasty. In Chapter 4, determinants predisposing to mortality and revision, or re-operation, will be explored using an advanced machine learning technique. In Chapter 7, machine learning algorithms to predict patient-reported outcome measures will be considered along with the key factors driving these predictions. Chapters 6 and 7 will explore the effect of implant fixation on revision rate due to periprosthetic fracture risk using ML and propensity score matching techniques with recommendations for optimal surgical practice for total hip replacement. As we will

explore in Chapter 4, cementless femoral implants are associated with a higher risk of revision and lower risk of mortality in comparison with cemented fixation. The cementless femoral replacement, considered in the second part of this thesis, is designed with the aim of addressing the problem of intraoperative periprosthetic fracture associated with cementless fixation. The studies in Chapters 8 and 9 are aimed to discern whether this objective is attained.

Chapter 4 Determinants of mortality, revision, and re-operation outcomes

4.1. Introduction

Chapter 3 of this thesis has highlighted the superior predictive ability exhibited by machine learning techniques compared to traditional statistical models, which may be attributed to algorithms' inherent capacity in intricate non-linear relationships. In previous studies traditional inference testing methods were employed to evaluate certain modifiable risk factors associated with increased early mortality risks following primary hip arthroplasty at *patient-level* such as advancing age, male gender, co-morbid conditions, and higher ASA grades (217, 241, 382), *surgical-level* such as cement type (217, 346), surgical approach (12), and diagnosis group (217) and *implant-level* such as prosthesis type, fixation, and articulation (12, 13). Similarly, statistical approaches using univariate and multivariate analyses were employed to analyse risk factors associated with increased risks of revision surgery. However, these studies inherently embark an incremental linear association assumption by utilising traditional statistical models. In addition, the selection of factors under investigation may introduce selection bias.

As outlined in Section 1.9, the risk factors pertaining to mortality, revision, and unfavourable functional outcomes in primary THR have been extensively investigated. Examination of SHAR and NJR annual reports reveals that the clinical decision-making, implant choice, and patient selection for primary elective hip replacement varies depending on a magnitude of factors, suggesting multilayer interactions among patient, surgical, and implant factors that predispose to safer outcomes (5, 6). Amongst these factors, there remains insufficient evidence about the importance of BMI, co-morbidities,

pre-operative diagnosis, and implant design, specifically fixation technique, on THR outcomes. Identification of these factors using machine learning techniques may provide an unbiased perspective for patient selection and better outcomes.

The aim of this chapter is to delineate the determinants of safety outcomes predicted in Chapter 3 and to address the issues of multicollinearity, overfitting, and factor selection by using the Cox proportional hazards model with elastic net regularisation (CoxNet algorithm) on the high-dimension SHAR and HES/NJR cohorts described in Chapter 2 and 3.

4.2. Patients and Methods

4.2.1. Data sources

This study uses the SHAR and HES/NJR linked datasets described previously. Chapter 2 provides a comprehensive description of the three data sources utilised in this study, including information on data collection, processing, and linkage. All patients who underwent primary elective hip arthroplasty in the SHAR extract between 2000 and 2018 (n=154,595) and the HES/NJR extract between 2009 and 2018 (n=355,933) were identified. Similar to Chapter 3, this study is in two parts; the first part uses the SHAR dataset to identify key factors driving safety and implant failure outcomes (mortality, revision and re-operation) followed by univariate and multivariate statistical analyses, and the second uses HES/NJR linked data to delineate the determinants for each of the mortality and revision outcomes. Table 4-1 represents the variables used in each dataset.

	SHAR		NJR		HES
Patient factors	Age		Age		Age
	Gender		Gender		Gender
	BMI		BMI		
	ASA grade		ASA grade		
	Side of operation		Side of operation		
	Hospital type		Unit type		Unit type
Surgical factors	Operation type		Operation type		Procedure type
	Diagnosis group		Diagnosis group		Length of stay
	Incision type		Incision type		
	Prothesis group		Prosthesis group		
	Cement type		Cement type		
Cup implant factors	Fixation		Fixation		
	Resurfacing		Resurfacing		
	Articulation		Articulation		
	Modular or Monoblock		Modular or Monoblock		
	Inner diameter		Inner diameter		
	Modification		Modification		
Femoral implant factors	Fixation		Fixation		
	Stem resurfacing		Stem resurfacing		
	Articulation		Articulation		
	Head size		Head size		
Outcomes CoxNet Survival Analysis	Death	1-year	Death (data linked with HES)		1-year
	Revision	2-year	Revision	2-year	
		5-year		5-year	
		10-year			
	Re-operation	2-year			
		5-year			
		10-year			

Table 4-1 Sources of data

The structured dataset used in the study comprised a total of twenty-one variables. These variables can be categorised into six patient-related variables, four surgical variables, and eleven implant-specific variables. The six patient variables included in the dataset are age, gender, body mass index (BMI), American Society of Anaesthesiologists (ASA) score, the side of the operation (left or right), and the type of hospital where the procedure was performed. The four surgical variables captured in the dataset are the diagnosis group

(indicating the specific condition or reason for the surgery), incision type (the type of surgical incision made), prosthesis group (categorising the type of hip prosthesis used), and cement type (the type of cement utilised in the procedure). The eleven implant-specific variables were further divided into six cup variables and five femoral variables. The cup variables included fixation type (how the cup component is secured to the bone), resurfacing (whether the cup is a total hip replacement or hip resurfacing), articulation type (the materials used for the bearing surfaces), modular or monoblock design (the structure of the implant), implant size, and the additional variable "modification" related to the cup. The femoral variables encompassed fixation type, resurfacing, articulation type, modular or monoblock design, and implant size.

4.2.2. Exposures

For both parts of the study, machine learning models were applied to forecast the outcomes of hip arthroplasty. Analyses were restricted to patients undergoing elective primary hip replacement. Patients who had undergone hip arthroplasty due to specific diagnoses such as trauma, tumour-related conditions, and degenerative conditions other than osteoarthritis were excluded. Patients below the age of 18 were excluded to minimise potential confounding factors that could arise from the different outcomes observed in paediatric patients (369-371).

4.2.3. Outcomes of interest

There were three primary outcome measures in the SHAR dataset: mortality, revision, and re-operation, and two primary outcome measures in the HES/NJR dataset: mortality and revision. Re-operation is defined as any additional surgery to the hip, regardless of the actions taken to any of the implant components (replaced, extracted or left untouched) whereas revision is defined as a re-operation where at least one component is exchanged, extracted from, or added, to the prosthesis. Machine learning survival analysis techniques

(CoxNet) were used with an analysis of time-to-event and a selection of key predictor variables that were assigned appropriate weights, whilst shrinking other variables' coefficients towards zero (see 'machine learning and statistical methods' below). For mortality, we considered 1-year survival analysis. For re-operations and revisions, we considered survival analyses for 2 years, 5 years and 10 years as separate outcomes of interest.

4.2.4. Machine learning and statistical methods

4.2.4.1. Descriptive statistics

Descriptive statistics were performed and described in Chapter 3 using Kaplan-Meier (KM) non-parametric estimates to describe mortality and implant survival probabilities over time across various age, sex groups, and by year in both datasets (372).

4.2.4.2. Model specification

Predictors of mortality and implant survival were examined using lasso penalised Cox proportional hazard regression (CoxNet). To compute survival models using approach, we used a combination of elastic net regularisation, through using lasso penalties for feature selections, and their corresponding coefficients for the predictors. Variable selection in CoxNet refers to the process of applying a penalty to the coefficients and shrinking some coefficients towards zero, or in other words, identifying a subset of features that are most relevant for predicting survival outcomes. Univariable hazard ratios, 95% confidence intervals, and the P values were generated to evaluate the overall effect of each CoxNet-selected predictor on each outcome endpoint. Multivariable models were fitted with further adjustments for age, gender and/or ASA to calculate the relative effect sizes of the machine learning selected predictors. As before, analyses were performed using R statistical computing environment version 4.3.0 R. Packages 'glmnet'

(version 4.1.7) and 'survival' (version 3.5.5) were used for computing the lasso penalised Cox proportional hazard regression (CoxNet) model.

4.3. Results

4.3.1. Swedish Hip Arthroplasty Register

Descriptive figures related to the SHAR survival outcome measures for death, revision, and re-operation are presented in Table 4-1.

Outcome	Time from primary operation	Derivation cohort (N=154,595) N	%
Mortality	1 year	1,301	0.84
Revision	2 years	3,011	1.95
	5 years	4,179	2.70
	10 years	5,133	3.32
Re-operation	2 years	4,428	2.86
	5 years	5,897	3.81
	10 years	7,045	4.56

Table 4-2 SHAR survival outcomes

4.3.1.1. Summary statistics of outcomes

Summary statistics, Kaplan-Meier (KM) estimates, and Hazard Ratios (HR) for cumulative incidence of mortality, revision, and re-operation outcomes, respectively, at each time point by age and sex, and by year are shown in Chapter 3 (Tables 3-3, 3-4, 3-5, 3-6, and Appendices A3.1 & A3.2).

4.3.1.2. Lasso penalised Cox proportional hazard regression (CoxNet model)

The elastic net regularisation, utilising lasso penalty terms, identified a set of eight variables (ten features) and their corresponding coefficients that are most relevant for predicting one-year mortality survival outcome. These features are categorised into patient, diagnostic, and implant factors, as illustrated in Table 4-3. Modifiable factors were comprised of implant-specific features, primarily dual mobility cups and cemented

femoral fixation, whilst intrinsic features included: age, male gender, ASA group, low Body Mass Index (BMI), private hospitals, and idiopathic necrosis/other diagnosis group. A positive coefficient indicates that the feature increases the risk of mortality whilst a negative coefficient, such as "Private Hospital", is protective.

	CoxNet-selected variables (n)	Features	One-year mortality coefficient (n = 8)
Patient factors	Age		+ 0.045
	Gender	Male	+ 0.259
	ASA		+ 0.820
	BMI	Underweight Group [0 - 18.5)	+ 0.340
	Hospital Type	Private	- 0.066
Surgical factors	Diagnosis	Idiopathic Necrosis	+ 1.033
		Other	+ 0.451
Implant factors	Cup Articulation/ Modularity	DMC-Monoblock	+ 0.234
		DMC-Modular	+ 4.065
	Femoral Fixation	Cemented	+ 0.040

Table 4-3 SHAR one-year mortality CoxNet model
selected variables and coefficients

The CoxNet-selected features for predicting revision survival outcomes at two-year, five-year and ten-year periods are shown in Table 4-4, alongside their coefficients. In total, eight variables (ten features) were selected for each of the two-year and five-year endpoints and eleven variables (fifteen features) for the 10-year revision outcome. Patient factors, such as age, showed minimal negative impact on revision risk and is likely a confounding factor whilst Male gender and higher ASA grades had positive coefficients, indicating an increased risk of revision at all time points. In contrast to the SHAR one-year mortality CoxNet-selected features, obesity and rural hospitals were possibly linked with

higher and protective revision risks, respectively. The diagnosis of idiopathic necrosis had positive coefficients across all time intervals, in line with the mortality survival outcome. Implant factors, such as cup resurfacing, metal cup articulation, and cementless femoral implant fixation, showed positive coefficients, indicating an increased risk of revision across the three revision endpoints.

	CoxNet-selected variables (n)	Features	Two-year revision coefficient (n = 8)	Five-year revision coefficient (n = 8)	Ten-year revision coefficient (n = 11)
Patient factors	Age				- 0.001
	Gender	Male	+ 0.34	+ 0.34	+ 0.30
	ASA		+ 0.24	+ 0.15	+ 0.11
	BMI	Obesity Class I [30 - 35)	+ 0.12	+ 0.09	+ 0.11
		Obesity Class II [35 - 40)	+ 0.42	+ 0.32	+ 0.32
		Obesity Class III ≥ 40	+ 0.77	+ 0.56	+ 0.49
	Hospital Type	Rural	- 0.001	- 0.006	- 0.04
Surgical factors	Diagnosis	Idiopathic Necrosis	+ 0.44	+ 0.42	+ 0.36
Implant factors	Cup Resurfacing	Resurfaced		+ 0.47	+ 0.92
	Cup Articulation	Metal (Standard)			+ 0.69
		Poly (X-link)			- 0.23
	Cup Modularity	Modular	+ 0.09		
	Cup Modification	Standard	- 0.03	- 0.13	- 0.15
		Liped			+ 0.010
	Femoral Fixation	Cementless	+ 0.18	+ 0.18	+ 0.22
	Femoral Articulation	Head Metal (Non-titanium)			+ 0.06

Table 4-4 SHAR revision CoxNet models

Table 4-5 illustrates the CoxNet-selected variables for re-operation survival outcomes at two-year, five-year, and ten-year periods with their corresponding coefficients. Overall, similar features to revision outcomes were identified. Inflammatory and secondary arthritis were linked with higher risks for re-operation at five and ten years. Similarly, a posterior approach had a negative coefficient indicating a lower risk of re-operation at all time points. Cementless femoral fixation had a higher risk of re-operation and comparable coefficients to revision outcomes. Cup implant factors, such as resurfacing and specific types of cup articulation, have varying effects on re-operation risk, with Poly (X-link) seeming to have a negative protective effect.

	CoxNet-selected variables (n)	Features	Two-year re-operation coefficient (n = 10)	Five-year re-operation coefficient (n = 11)	Ten-year re-operation coefficient (n = 14)
Patient factors	Age				- 0.001
	Gender	Male	+ 0.38	+ 0.37	+ 0.34
	ASA		+ 0.30	+ 0.22	+ 0.18
	BMI	Obesity Class I [30 - 35)	+ 0.26	+ 0.21	+ 0.20
		Obesity Class II [35 - 40)	+ 0.60	+ 0.48	+ 0.45
		Obesity Class III ≥ 40	+ 0.86	+ 0.67	+ 0.56
	Hospital Type	County	+ 0.04	+ 0.008	+ 0.04
		Rural	- 0.16	- 0.16	- 0.15
Surgical factors	Diagnosis	Idiopathic Necrosis	+ 0.59	+ 0.56	+ 0.26
		Other Secondary Arthrosis	+ 0.10	+ 0.22	+ 0.48
		Inflammatory		+ 0.09	+ 0.27
	Incision Type	Direct Lateral	+ 0.005		
		Posterior	- 0.15	- 0.11	- 0.09
Cup implant factors	Resurfacing	Resurfaced		+ 0.51	+ 1.19
	Size	Inner Diameter			- 0.01
	Articulation	DMC Monoblock	+ 0.23	+ 0.02	
		DMC Modular	+ 1.39	+ 0.54	+ 0.31
		Metal (Standard)		+ 0.21	+ 1.30
		Poly (X-link)		- 0.10	- 0.27
	Modification	Standard	- 0.04	- 0.08	- 0.08
		Liped			+ 0.07
Femoral factors	Fixation	Cementless	+ 0.11	+ 0.15	+ 0.20
	Articulation	Head Metal (Non-titanium)	+ 0.06	+ 0.06	+ 0.13

Table 4-5 SHAR re-operation CoxNet models

4.3.1.3. Factors affecting safety

Univariable and multivariable one-year mortality survival analyses based on CoxNet-selected variables are displayed in 4-6 and 4-7, respectively. Primary hip arthroplasty procedures were safer in younger patients, females, those with less co-morbidities (lower ASA grade), and patients with a BMI of 18.5 or above. There was a strong trend towards superior survival in the independent sector when compared to public hospitals.

CoxNet-selected variables		Univariable analysis (non-adjusted)		Univariable analysis (adjusted for age and gender)		Univariable analysis (adjusted for age, gender and ASA)	
		Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
Age		1.08 (1.07-1.08)	***	1.08 (1.07-1.08)	***	1.06 (1.05-1.06)	***
Gender	Male	1.44 (1.29-1.60)	***	1.72 (1.54-1.92)	***	1.58 (1.41-1.76)	***
ASA		3.37 (3.10-3.66)	***	2.62 (2.40-2.87)	***	2.62 (2.40-2.87)	***
BMI	Underweight [0 - 18.5)	3.28 (2.31-4.66)	***	2.83 (1.98-4.03)	***	2.59 (1.82-3.70)	***
Hospital type	Private	0.48 (0.40-0.56)	***	0.57 (0.48-0.67)	***	0.65 (0.55-0.77)	***
Diagnosis	Idiopathic Necrosis	4.23 (3.51-5.10)	***	4.27 (3.54-5.15)	***	3.50 (2.90-4.23)	***
	Other	5.71 (2.56-12.75)	***	6.08 (2.72-13.56)	***	4.09 (1.83-9.14)	***
Cup Articulation/ Modularity	DMC Monoblock	4.17 (3.07-5.65)	***	2.87 (2.12-3.90)	***	2.08 (1.53-2.83)	***
	DMC Modular	57.41 (14.34-229.8)	***	274.5 (68.18-1105.91)	***	112.47 (27.89-453.48)	***
Femoral Fixation	Cemented	2.84 (2.44-3.30)	***	1.36 (1.15-1.61)	***	1.37 (1.16-1.62)	***

Table 4-6 SHAR one-year mortality univariate analysis using CoxNet-selected variables
*** p<0.05, ** p<0.01, *** p<0.001**

A pre-operative diagnosis of idiopathic necrosis or "other", which includes cases other than degenerative osteoarthritis, inflammatory or secondary arthritis, is associated with a higher risk of mortality. The choice of cemented femoral fixation was significantly correlated with an increased risk of mortality in comparison to cementless seating. Similarly, the use of dual mobility cups (DMC), whether modular or monoblock, had a

higher risk of death. However, the wide confidence interval for DMC modular (n = 5, Table 2-1) cups reduces the precision of the findings.

CoxNet-selected variables		Multivariable analysis (non-adjusted)		Multivariable analysis (adjusted for age and gender)		Multivariable analysis (adjusted for age, gender and ASA)	
		Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
Age		1.05 (1.04-1.06)	***	--	--	--	--
Gender	Male	1.67 (1.50-1.87)	***	--	--	--	--
ASA		2.44 (2.23-2.67)	***	2.42 (2.21-2.64)	***	--	--
BMI	Underweight [0 - 18.5)	2.22 (1.55-3.17)	***	2.18 (1.52-3.13)	***	2.10 (1.46-3.04)	***
Hospital type	Private	0.70 (0.59-0.83)	***	0.71 (0.60-0.84)	***	0.73 (0.62-0.87)	***
Diagnosis	Idiopathic Necrosis	3.17 (2.61-3.84)	***	2.93 (2.41-3.56)	***	2.89 (2.37-3.54)	***
	Other	4.08 (1.82-9.13)	***	3.24 (1.43-7.34)	**	3.43 (1.50-7.86)	**
Cup Articulation/ Modularity	DMC Monoblock	1.48 (1.08-2.02)	*	1.41 (1.03-1.94)	*	1.38 (1.00-1.92)	*
	DMC Modular	130.37 (32.26-526.74)	***	82.46 (17.40-390.64)	***	58.24 (9.95-340.72)	***
Femoral Fixation	Cemented	1.36 (1.15-1.60)	***	1.44 (1.21-1.71)	***	1.44 (1.20-1.72)	***

Table 4-7 SHAR one-year mortality multivariate analysis using CoxNet-selected variables
*** p<0.05, ** p<0.01, *** p<0.001**

4.3.1.4. Factors affecting implant survival

Multivariable two- and five-year revision and re-operation survival analyses are displayed in 4-8 and 4-9, respectively. Unlike mortality, age was not selected by the CoxNet model whilst obesity increased the risk of revision and re-operation at two and five years. Female gender, lower ASA grades and procedures undertaken in rural hospitals had a statistically significant lower risk of revision and re-operation at two and five years.

A pre-operative diagnosis of idiopathic necrosis was significantly correlated with elevated risks of revision and re-operation, whereas the diagnosis of secondary arthritis exhibited an increased risk of re-operation, while inflammatory arthritis was found to elevate the risk of re-operation specifically at the five-year interval only.

CoxNet-selected variables		Two-year revision Multivariable analysis (non-adjusted)		Two-year revision Multivariable analysis (adjusted for age and gender)		Five-year revision Multivariable analysis (non-adjusted)		Five-year revision Multivariable analysis (adjusted for age and gender)	
		HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Gender	Male	1.65 (1.53-1.77)	***	--	--	1.61 (1.51-1.71)	***	--	--
ASA		1.41 (1.33-1.49)	***	1.32 (1.24-1.40)	***	1.30 (1.23-1.36)	***	1.27 (1.21-1.34)	***
BMI	Obesity Class I [30 - 35)	1.42 (1.30-1.54)	***	1.47 (1.35-1.61)	***	1.35 (1.25-1.45)	***	1.36 (1.26-1.46)	***
	Obesity Class II [35 - 40)	2.02 (1.78-2.29)	***	2.19 (1.92-2.49)	***	1.80 (1.25-1.45)	***	1.84 (1.64-2.06)	***
	Obesity Class III ≥ 40	3.08 (2.50-3.80)	***	3.46 (2.80-4.28)	***	2.59 (2.13-3.14)	***	2.68 (2.20-3.27)	***
Hospital type	Rural	0.85 (0.78-0.92)	***	0.85 (0.79-0.92)	***	0.86 (0.80-0.92)	***	0.86 (0.80-0.92)	***
Diagnosis	Idiopathic Necrosis	2.03 (1.73-2.39)	***	2.14 (1.82-2.52)	***	1.99 (1.73-2.28)	***	2.07 (1.79-2.38)	***
Cup Resurfacing	Resurfaced					2.38 (1.95-2.91)	***	2.63 (2.14-3.24)	***
Cup Modularity	Monoblock	1.08 (0.94-1.25)	/	1.09 (0.95-1.26)	/				
Cup Modification	Standard	0.87 (0.77-0.99)	*	0.84 (0.74-0.95)	**	0.82 (0.76-0.88)	***	0.81 (0.75-0.87)	***
Femoral Fixation	Cementless	1.32 (1.20-1.45)	***	1.52 (1.37-1.68)	***	1.34 (1.24-1.44)	***	1.41 (1.30-1.53)	***

Table 4-8 SHAR two-year and five-year revision multivariate analysis using CoxNet-selected variables
*** p<0.05, ** p<0.01, *** p<0.001**

Within the realm of implant-specific factors, the use of a standard cup modification decreased the risk of implant failure. While the use of a metal cup heightened the probability of requiring a re-operation after five years, poly x-link material substantially reduced this likelihood. Cemented femoral fixation demonstrated favourable effects by reducing the implant failure risks whilst hip resurfacing was identified as a contributor to increased revision and re-operation risks at the five-year mark.

CoxNet-selected variables		Two-year re-operation Multivariable analysis (non-adjusted)		Two-year re-operation Multivariable analysis (adjusted for age and gender)		Five-year re-operation Multivariable analysis (non-adjusted)		Five-year re-operation Multivariable analysis (adjusted for age and gender)	
		HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Gender	Male	1.61 (1.52-1.71)	***	--	--	1.58 (1.50-1.66)	***	--	--
ASA		1.42 (1.35-1.48)	***	1.35 (1.28-1.42)	***	1.31 (1.25-1.36)	***	1.28 (1.23-1.34)	***
BMI	Obesity Class I [30 - 35)	1.48 (1.38-1.59)	***	1.53 (1.42-1.64)	***	1.41 (1.32-1.50)	***	1.42 (1.33-1.51)	***
	Obesity Class II [35 - 40)	2.12 (1.91-2.35)	***	2.25 (2.02-2.50)	***	1.89 (1.72-2.08)	***	1.92 (1.75-2.12)	***
	Obesity Class III ≥ 40	2.92 (2.44-3.49)	***	3.22 (2.68-3.87)	***	2.45 (2.07-2.91)	***	2.56 (2.15-3.04)	***
Hospital Type	County	1.14 (1.06-1.23)	***	1.12 (1.04-1.21)	**	1.08 (1.02-1.16)	**	1.07 (1.01-1.15)	*
	Rural	0.82 (0.76-0.89)	***	0.82 (0.76-0.88)	***	0.81 (0.76-0.87)	***	0.81 (0.76-0.87)	***
Diagnosis	Idiopathic Necrosis	2.07 (1.81-2.36)	***	2.12 (1.85-2.43)	***	2.03 (1.80-2.28)	***	2.07 (1.83-2.33)	***
	Other Secondary Arthrosis	1.39 (1.19-1.63)	***	1.47 (1.25-1.72)	***	1.49 (1.31-1.69)	***	1.52 (1.33-1.73)	***
	Inflammatory					1.51 (1.25-1.82)	***	1.58 (1.31-1.91)	***
Incision Type	Direct Lateral	0.98 (0.61-1.57)	/	1.03 (0.64-1.64)	/				
	Posterior	0.75 (0.47-1.91)	/	0.78 (0.49-1.24)	/	0.79 (0.75-0.83)	***	0.79 (0.75-0.83)	***
Cup Resurfacing	Resurfaced					2.13 (1.77-2.57)	***	2.26 (1.86-2.74)	***
Cup Articulation	DMC Monoblock	1.58 (1.25-1.99)	***	1.43 (1.13-1.82)	**	1.24 (1.00-1.54)	*	1.17 (0.94-1.46)	/
	DMC Modular	7.73 (1.92-31.13)	**	NA	NA	5.25 (1.30-21.13)	*	5.77 (1.40-23.77)	*
	Metal (Standard)					2.01 (1.33-3.05)	***	1.96 (1.29-2.97)	**
	Poly (X-link)					0.79 (0.75-0.84)	***	0.79 (0.74-0.84)	***
Cup Modification	Standard	0.92 (0.86-0.99)	*	0.89 (0.83-0.96)	**	0.92 (0.86-0.98)	*	0.91 (0.85-0.97)	**
Femoral Fixation	Cementless	1.22 (1.13-1.31)	***	1.36 (1.25-1.47)	***	1.31 (1.23-1.39)	***	1.37 (1.28-1.47)	***
Femoral Articulation	Head Metal (Non-titanium)	1.24 (1.13-1.35)	***	1.18 (1.08-1.29)	***	1.23 (1.15-1.33)	***	1.22 (1.13-1.31)	***

Table 4-9 SHAR two-year and five-year re-operation multivariate analysis using CoxNet-selected variables * p<0.05, ** p<0.01, *** p<0.001

4.3.2. National Joint Registry

Descriptive figures related to the NJR survival outcome measures for death and revision are presented in Table 4-10.

Outcome	Time from primary operation	NJR Derivation cohort (N=355,933) N	%
Mortality	30 days	506	0.14
	60 days	782	0.21
	90 days	1,047	0.29
	1 year	3,636	1.02
Revision	6 months	1,824	0.51
	1 year	2,450	0.68
	2 years	3,362	0.94
	5 years	4,671	1.31

Table 4-10 NJR survival outcomes

4.3.2.1. Summary statistics of outcomes

Summary statistics, Kaplan-Meier (KM) estimates, and Hazard Ratios (HR) for the cumulative incidence of mortality, and revision outcomes, respectively, and at each time point by age and sex, and by year, are shown in Chapter 3 (Tables 3-9, 3-10, 3-11, and 3-12).

4.3.2.2. Lasso penalised Cox proportional hazard regression (CoxNet model)

The CoxNet model identified a set of ten variables (thirteen features) and their corresponding coefficients that are most relevant for predicting one-year mortality survival outcome. Table 4-11 categorises these features into patient, diagnostic, and implant factors. Modifiable implant-specific features primarily included fixation and cup and femoral articulation, whilst patient features comprised of age, male gender, ASA group, low Body Mass Index (BMI), non-private hospitals and idiopathic necrosis/other diagnosis group.

	CoxNet-selected variables (n)	Features	One-year mortality coefficient (n = 10)
Patient factors	Age		+ 0.053
	Gender	Male	+ 0.390
	ASA		+ 0.894
	BMI	Underweight Group [0 - 18.5)	+ 0.230
	Hospital Type	Public	+ 0.049
		University	+ 0.062
Surgical factors	Diagnosis	Idiopathic Necrosis	+ 0.506
		Primary arthrosis	- 0.155
		Other	+ 1.204
	Incision Type	Other	- 0.031
Implant factors	Cup Articulation	DMC Modular	+ 0.172
	Cup Fixation	Cementless	- 0.038
	Femoral Articulation	Metal	+ 0.100

Table 4-11 NJR mortality CoxNet models

The CoxNet-selected variables for revision survival outcomes at two-year and five-year periods are shown in Table 4-12. Seven variables were selected for the two-year revision mark (twelve features) and five variables were CoxNet-selected for the five-year revision endpoints (six features). Age showed a minimal negative impact on the five-year revision risk and is likely a confounding factor, whilst Male gender and higher ASA grades had positive coefficients, indicating an increased two-year revision risk. In contrast, in the one-year mortality features, obesity was shown to increase revision risk. The diagnosis of idiopathic necrosis had a positive two-year coefficient, in line with the mortality survival outcome. Implant factors including cementless femoral implant fixation, monoblock cups,

and metal cup articulation showed positive coefficients, indicating an increased risk of revision as highlighted in Table 4-12.

	CoxNet-selected variables (n)	Features	Two-year revision coefficient (n = 7)	Five-year revision coefficient (n = 5)
Patient factors	Age			- 0.006
	Gender	Male	+ 0.091	
	ASA		+ 0.050	
	BMI	Healthy weight [18.5 - 25)	- 0.027	
		Overweight [25 - 30)	- 0.011	- 0.007
		Obesity Class I [30 - 35)	Ref.	Ref.
		Obesity Class II [35 - 40)	+ 0.196	
		Obesity Class III ≥ 40	+ 0.323	+ 0.009
Surgical factors	Diagnosis	Idiopathic Necrosis	+ 0.053	
		Primary arthrosis	- 0.172	
		Other secondary arthrosis	+ 0.662	
	Incision Type	Posterior	- 0.006	
	Prosthesis Group	Cementless		+ 0.013
Implant factors	Cup Articulation	Metal (Standard)		+ 0.641
	Cup Modularity	Monoblock	+ 0.125	
	Femoral Fixation	Cementless	+ 0.112	+ 0.070

Table 4-12 NJR revision CoxNet models

4.3.2.3. Factors affecting safety

Multivariable one-year mortality survival analyses based on CoxNet-selected variables are displayed in 4-13. Primary hip arthroplasty procedures in the NJR were safer in younger patients, females, those with lower ASA grades and patients with a BMI of 18.5

or above. There was a strong trend towards inferior survival in public and university hospitals.

CoxNet-selected variables		Multivariable analysis (non-adjusted)		Multivariable analysis (adjusted for age and gender)	
		Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
Age		1.05 (1.05-1.06)	***	--	--
Gender	Male	1.65 (1.54-1.77)	***	--	--
ASA		2.46 (2.32-2.6)	***	2.43 (2.30-2.58)	***
BMI	Underweight [0 - 18.5)	1.38 (1.28-1.49)	***	1.34 (1.24-1.45)	***
Hospital type	Public	1.43 (1.28-1.59)	***	1.41 (1.27-1.57)	***
	University	1.49 (1.32-1.69)	***	1.47 (1.30-1.67)	***
Diagnosis	Idiopathic Necrosis	1.85 (1.38-2.50)	***	1.89 (1.40-2.54)	***
	Primary arthrosis	0.84 (0.64-1.10)		0.91 (0.70-1.19)	
	Other	3.58 (2.62-4.90)	***	3.74 (2.73-5.13)	***
Cup Articulation/ Modularity	DMC Monoblock	1.44 (1.07-1.94)	*	1.38 (1.02-1.87)	*
Cup Fixation	Cementless	0.92 (0.85-0.99)	*	0.92 (0.86-0.99)	*
Femoral Articulation	Metal	1.21 (1.10-1.34)	***	1.28 (1.16-1.42)	***

Table 4-13 NJR one-year mortality multivariate analysis without and with adjustment to age and gender
*** p<0.05, ** p<0.01, *** p<0.001**

A pre-operative diagnosis of idiopathic necrosis or "other" were associated with a higher mortality risk. The choice of cementless cup fixation significantly correlated with a decreased risk of mortality in comparison to cemented fixation. Similarly, the use of dual mobility monoblock cups (DMC) or metal femoral heads had a higher risk of death.

4.3.2.4. Factors affecting implant survival

The multivariable survival analyses for revisions at both the two-year and five-year time points are presented in Table 4-14.

CoxNet-selected variables		Two-year revision Multivariable analysis (non-adjusted)		Two-year revision Multivariable analysis (adjusted for age and gender)		Five-year revision Multivariable analysis (non-adjusted)		Five-year revision Multivariable analysis (adjusted for age and gender)	
		Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
Age				--	--	0.98 (0.98-0.99)	***	--	--
Gender	Male	1.24 (1.15-1.34)	***	--	--				
ASA		1.15 (1.07-1.23)	***	1.15 (1.07-1.24)	***				
BMI	Overweight [25-30)	0.81 (0.74-0.90)	***	0.81 (0.74-0.89)	***	0.82 (0.75-0.90)	***	0.81 (0.74-0.89)	***
	Obesity Class I [30 - 35)	Ref.		Ref.		Ref.		Ref.	
	Obesity Class II [35 - 40)	1.29 (1.14-1.45)	***	1.28 (1.14-1.45)	***				
	Obesity Class III ≥ 40	1.55 (1.31-1.83)	***	1.53 (1.30-1.82)	***	1.36 (1.14-1.64)	***	1.40 (1.17-1.68)	***
Diagnosis	Idiopathic Necrosis	1.28 (0.98-1.68)		1.27 (0.97-1.67)					
	Primary arthrosis	0.72 (0.60-0.88)	**	0.76 (0.62-0.92)	**				
	Other Secondary Arthrosis	3.26 (1.87-5.69)	***	3.26 (1.87-5.69)	***				
Incision Type	Posterior	0.86 (0.79-0.93)	***	0.86 (0.79-0.93)	***				
Prosthesis Group	Cementless					1.04 (0.81-1.34)		1.05 (0.82-1.34)	
Cup Articulation	Metal					2.42 (1.98-2.96)	***	2.40 (1.96-2.93)	***
Femoral Fixation	Cementless	1.35 (1.25-1.46)	***	1.35 (1.25-1.47)	***	1.19 (0.93-1.53)		1.20 (0.93-1.53)	

Table 4-14 NJR two-year and five-year revision multivariate analysis using without and with adjustment to age and gender * p<0.05, ** p<0.01, *** p<0.001

In contrast to its effect on mortality, obesity was found to elevate the risk of revision at both the two-year and five-year junctures. In contrast, female gender and lower ASA grades were associated with a statistically significant reduction in the risk of revision at

the two-year interval. On the other hand, a diagnosis of idiopathic necrosis was not associated with increased revision risks, whereas the diagnosis of secondary arthritis exhibited an increased revision risk at the two-year interval only. In terms of implant-specific factors, the use of cementless femoral fixation is significantly associated with a heightened risk of revision within the first two years postoperatively, albeit not at the five-year mark. Conversely, metal cups exhibit detrimental effects by diminishing the survival of implants over a five-year duration.

4.4. Discussion

This study introduced a combined survival machine learning approach and multivariate analyses to identify and examine the factors predicting mortality and implant failure outcomes following primary elective total hip replacement. The outcomes of this study can be used to assist in the selection of patients and implant designs for better surgical outcomes.

4.4.1. Patient factors

Increasing age has been identified to be associated with a significant, albeit small, increase in mortality risk and a decrease in long-term revision risk in the two registries. Similarly, males exhibited higher risks for both outcomes. Although no adjustments for co-morbidities were conducted in this study, these findings are in agreement with the KM estimates described in Chapter 3 and with previous studies, suggesting a possible physiological aging process in the elderly and the likelihood of surpassing the lifetime risk in younger individuals (see Section 1.9 and (226-228)). Moreover, the existing evidence supports the absence of a significant association between gender and mortality or revision outcomes following primary total hip arthroplasty (227, 283-285). Therefore, age and gender differences were adjusted for in our multivariate analyses.

Co-morbidities, as measured by the ASA grade, increased the incidences of mortality and implant failure at all time points, and were more pronounced in the former outcome. Following adjustments to age and gender, higher ASA grades increased mortality risks by almost 2.5 times (HR_{SHAR} 2.42, HR_{NJR} 2.43, $p < 0.001$) and two-year implant failure risk by approximately 15% (in the NJR; $HR_{revision}$ 1.15) and 32 – 35% (in the SHAR; $HR_{revision}$ 1.32, $HR_{reop.}$ 1.35). Whilst this aligns with the existing moderate evidence suggesting an elevated risk of revision associated with higher co-morbidity indices (see Section 1.9), it underscores the importance of prospective clinical trials in this domain.

This study furnishes substantiation of an augmented mortality hazard within the underweight patient cohort, alongside a heightened risk of revision surgery observed among the obese patient group. This is in concurrence with prior research reporting similar obesity paradox patterns (see Section 1.9 and (12, 243, 244)). Whilst our study is inherently limited to suggest restriction of intervention solely based on an individual's BMI, it is prudent to consider the trade-off between these two factors.

4.4.2. Surgical factors

Private healthcare provision was shown in both registries to have superior safety outcomes compared to public hospitals, whilst rural hospitals in the SHAR, but not in the NJR, were associated with lower incidences of re-do procedures. Examination of the SHAR annual report alludes to the increasing proportion of procedures undertaken at private healthcare facilities with the majority of private patients having lower ASA grades and a pre-operative diagnosis of primary osteoarthritis (5). Likewise, the posterior surgical approach in the NJR was found to reduce two-year revision risk by approximately 14%, concurring with previous studies (293, 311), however, this was not pronounced at the five-year period nor in the SHAR dataset. Overall, it is important to note that neither of our extracted data included information pertaining to unit volume or surgical expertise,

therefore, such associations and their significance are likely attributed to selection bias and should be interpreted with caution.

In terms of surgical indication, a pre-operative diagnosis of non-traumatic idiopathic necrosis approximately doubled the risk of death in both registries. Whilst this effect persisted, alongside inflammatory arthritis for re-operation risk, with statistical significance for implant failure outcomes in the SHAR ($p < 0.001$), it was not statistically significant in the NJR. Although there is limited recent evidence focusing on the effect of avascular necrosis on mortality and revision outcomes, a previous SHAR study, of 427,806 observations between 1995 and 2011 reported heightened revision risk in patients with non-traumatic femoral head necrosis, particularly in the first two-years postoperatively (383). The evidence, however, remains disputed with earlier clinical studies having reported high failure rates for this patient group (384, 385), whereas other observational studies highlighted comparable revision outcomes to the primary osteoarthritis group (386-388). The discrepancy between these studies was thought to be primarily attributed to the varying durations of follow-up as well as potential implant selection bias or insufficient sample sizes (383).

4.4.3. Implant factors

Metal articulation was found to increase the one-year mortality risk associated with the use of metal heads by 28%, and the five-year revision risk associated with the use of metal cups by 2.4 times in the NJR. Whilst in the SHAR, only hip resurfacing arthroplasty was found to increase the five-year revision risk by 2.6 times. Previous studies emphasised the adverse events associated with metal-on-metal wear, local tissue reactions, and the elevated metal ion levels, which all potentially contribute to the inferior outcomes for metal articulations (320, 321). Additionally, there are several factors that might confound the association between hip resurfacing arthroplasty and increased five-year revision risk

in the SHAR: the limitations of using revision as an end point (see Section 1.8); the low proportion of HRA in Sweden with no procedures performed in 2022 (5); limited adjustment for patient activity levels pre- and post-operatively; unit and surgeon volume and expertise; and patient selection.

Cemented fixation and dual mobility cups were found to have a statistically significant increase in mortality risk in both registries whilst the NJR identified metal femoral heads as having a similar effect. However, the association between dual mobility cups and the outcome is likely of a confounding nature and could be influenced by the characteristics of the patient population who receive these implants (6). Our adjusted multivariate analyses highlight that cemented fixation may increase the one-year mortality risk by 9% (NJR, cemented cups) and 44% (SHAR, cemented stems), whilst lowering the implant failure risk by 35% (NJR, uncemented stems) and 52% (SHAR, uncemented stems). The discrepancy between the two registries in the identified component-related (stem versus cup) mortality risk adds an additional layer of complexity. Garland *et al.* in a SHAR matched cohort study, whilst not reaching statistical significance, reported an increased risk of death in the hybrid group (cemented femoral component), proposing that cemented femoral fixation is likely associated with higher incidences of thromboembolic events (348). Similarly, Trela-Larsen *et al.* applied a flexible parametric survival regression for personalised estimation of one-year mortality risk after elective hip and knee arthroplasty using the NJR as a training platform and externally validated their algorithms using the Norwegian Arthroplasty Register (NAR) (378). The authors in their supplementary materials reported similar findings for the Norwegian Arthroplasty Registry (NAR; n = 60,814, HR 1.43 95% CI 1.01 to 2.02, p = 0.04), but not the National Joint Registry for England and Wales (NJR; n = 340,033, HR 1.06, 95% CI 0.96 to 1.17, p = 0.25). While this holds for our SHAR findings, it contrasts with the NJR, suggesting the need for

further investigations. It is noteworthy, however, that our study did not investigate for specific reasons for mortality or revision risks. Interpreting the findings considering existing literature (expounded in Section 1.11), the risk-benefit assessment of fixation type in primary THR needs to account for multiple co-variates to address issues related to selection bias and confounding by indication. Chapter 7 will explore the effect of implant fixation on safety and implant failure outcomes using a propensity score matching technique.

4.4.4. Findings in context

The literature on risk factors of death and implant failure in primary THR is highly debated and is covered in-depth in Sections 1.9 and 1.10. This study supports the approach of López-López *et al,* in highlighting that a combination of factors, in their study [metal-on-metal, small femoral head sizes and cemented fixation], are associated with varying outcome endpoints (322). Our study underscores the effects of different factors documented in previous literature, and metaphorically projects the significance of Malcolm Gladwell's idea of "The Law of the Few", wherein changes (even if relatively small) in patient or implant selection can have a disproportionate impact and reach "The Tipping Point" in terms of mortality and revision outcomes. Of note, our study highlighted the varying effects of patient's preoperative BMI (the trade-off between underweight and obesity), diagnosis (non-traumatic idiopathic necrosis), and co-morbidities (higher ASA grades) as well as implant factors specifically fixation type (cemented versus cementless fixations) and articulation surfaces (metal heads or cups).

4.4.5. Strengths and limitations

The strengths of this study include the large study population using two of the largest hip registries coupled with the use of machine learning survival techniques to select key predictors of outcomes. This enables an unbiased variable selection and adjustment for

potential confounding factors. The agreement in findings between the two registries further strengthens our confidence in the precision of measurements and the reliability of our adjusted multivariate analyses, suggesting that they may be generalised at a population level.

Limitations include, similar to Chapter 3, the observational nature of the study underpinned by unmeasured biases of registry data, and limited specifics on the severity of co-morbidities and the reasons for death or implant failure. The large amount of missing data in both registries, particularly incomplete ASA grades in the SHAR prior to year 2008 and BMI in the NJR, and our assumption of missing completely at random by eliminating missing observations rather than imputing variables, may have introduced bias to the findings. However, imputation introduces uncertainty and may result in biased estimates, which our methodology aimed to avoid (389). Similarly, our study lacked information related to patient-reported outcomes and thus falls short of adequately addressing the primary goal of primary elective total hip replacement, which is improving patients' quality of life and physical function. This is especially crucial when evaluating the effectiveness of interventions such as hip resurfacing arthroplasty.

4.4.6. Conclusions

This study has identified significant patient, surgical, and implant design factors predicting mortality and implant failure outcomes in primary elective total hip arthroplasty. Notably, the trade-off between augmented mortality incidence in underweight patients and heightened revision risk in obese individuals is to be considered. Particular caution should be exercised in patients with a pre-operative diagnosis of non-traumatic femoral head necrosis and those with higher co-morbidities to mitigate unfavourable outcomes.

Femoral fixation stands out as a significant modifiable prognostic factor. In this study, cemented fixation demonstrated superior implant survivability but higher mortality rates. Whilst this finding supports prior literature, further studies are needed to address issues related to selection bias, confounding by indication, and patient-reported outcome measures (PROMs). In Chapter 5, predictive models for PROMs following primary THR will be explored, whilst Chapters 6 and 7 will investigate the risk factors of periprosthetic femoral fractures necessitating revision surgery.

Chapter 5 Predicting quality of life measures following primary hip arthroplasty

5.1. Introduction

Mortality and revision outcomes are the most commonly reported variables following arthroplasty, initially collected by joint registries for the purpose of implant safety surveillance. However, these do not reflect the ultimate outcome for patients (limitations related to revision as endpoint outcome are discussed in Section 1.8). On the other hand, pain and poor health-related quality of life (HRQoL) measures serve as the primary indication for primary elective total hip arthroplasty (181). The efforts undertaken by hip joint registries to monitor outcomes related to THR culminated in the adoption of HRQoL and disease-specific patient-reported outcome measures (PROMs) collection, commenced in 2002 in the SHAR and 2009 in the NJR for England and Wales, and underscore their commitment to address patient experience. Whilst previous studies examined the differences in these measures (181, 390, 391), it remains challenging to predict whether THR delivers the patient-anticipated improvements due to variations in baseline health status, quality of life, and daily activities. Forecasting this exerts a multi-tiered impact, benefitting individuals by facilitating shared decision making, enhancing care delivery at the local level, and informing national policy and economic evaluations.

The European HRQoL 5-Dimensions (EQ-5D), despite conceptual differences compared to joint-specific PROMs, such as the Oxford Hip Score (OHS) and Harris Hip Score (HHS), uses standardised questions to comprehensively assess individuals' physical and mental wellbeing, and pain and functional levels across five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The EQ-5D index represents a global

measure of the HQoL, calculated based on specific weight value set, whilst a visual analogue scale (EQ-VAS) is separately recorded for general health (392). This tool was shown to have concurrent validity, acceptable ceiling effects, and high discriminatory power in several patient groups including THR (393-396). See Section 1.8.3 for further information regarding the measures of functional outcome in THR.

A recent systematic review sought to discern the predictors of EQ-5D and EQ-VAS changes among individual patients undergoing THR with findings indicative of limited or inconclusive evidence in this regard (397). Two studies constructed machine learning models to forecast changes in PROMs (EQ-5D index and EQ-VAS) following hip arthroplasty, however, despite achieving a good binary discriminative power, they were constrained either by their limited and non-representative sample size or the absence of relevant clinical data (398, 399).

The aim of this chapter is to develop machine learning models that predict meaningful improvements in one-year HRQoL and compare their predictive ability to traditional statistical linear models, using a large dataset from the Swedish Hip Arthroplasty Register.

5.2. Patients and Methods

5.2.1. Data sources

This study uses the SHAR dataset described previously in Chapter 2. All patients who underwent primary elective hip arthroplasty in the SHAR extract between 2008 and 2018 (n=154,595) were included. Observations characterised by the absence of recorded pre-operative and post-operative PROMs data (Charlney classification, EQ-5D index, EQ-VAS) were omitted from the dataset, resulting in a final dataset of 82,526 THRs. This study is in two parts; the first part uses the SHAR dataset to develop and internally validate the

machine learning algorithms in predicting one-year clinically important improvements in EQ-5D index, EQ-VAS and combined EQ-5D/EQ-VAS scores. This is followed by a comparison of the essential features influencing these predictions in the best performing machine learning model, coupled with multivariate analyses of these factors. Table 6-1 provides the variables used in the SHAR dataset.

	SHAR
Patient factors	Age
	Gender
	BMI
	ASA grade
	Side of operation
	Hospital type
Surgical factors	Operation type
	Diagnosis group
	Incision type
	Prothesis group
	Cement type
Cup implant factors	Fixation
	Resurfacing
	Articulation
	Modular or Monoblock
	Inner diameter
	Modification
Femoral implant factors	Fixation
	Stem resurfacing
	Articulation
	Head size
Pre-operative PROMs	EQ-5D
	EQ-VAS
	Charnley classification
One-year Outcomes	EQ-5D, SRM threshold
	EQ-5D, Paired threshold
	EQ-VAS
	Combined EQ-5D (SRM)/EQ-VAS
	Combined EQ-5D (Paired)/EQ-VAS

Table 5-1 Sources of data

5.2.2. Exposures (from previous chapter)

For both parts of the study, machine learning models were applied to forecast the outcomes of hip arthroplasty. Analyses were restricted to patients undergoing elective primary hip replacement. Patients who had undergone hip arthroplasty due to specific diagnoses such as trauma, tumour-related conditions, and degenerative conditions other than osteoarthritis were excluded. Patients below the age of 18 were excluded to minimise potential confounding factors that could arise from the different outcomes observed in paediatric patients (369-371).

5.2.3. Outcome of interest

The primary outcome measure in the SHAR dataset was one-year improvements in the EQ-5D index, EQ-VAS and combined EQ-5D/EQ-VAS scores, as set by the minimal clinically important difference (MCID) thresholds described below in sub-Section 5.2.3.4. More details pertaining to the data collection process are found in Chapter 2.

Broadly speaking, there are two variations of the EQ-5D questionnaire: a three-level (3L) version and the newer five-level (5L) release, which added two additional layers of severity response levels in attempts to address the ceiling effects, the low sensitivity, and enhance the discriminatory power of the 3L version (395, 400, 401). The SHAR adopted PROMs data collection using the 3L version up until 2017, at which point it transitioned to the 5L version. Eneqvist *et al.* compared the two versions in Western Sweden in terms of redistribution of responses, ceiling and floor effects, and correlations of EQ-VAS and EQ-5D indices (402). The authors reported a good agreement for EQ-VAS (Spearman's rank correlation coefficient, r = 0.71 pre-operatively, Spearman r = 0.87 post-operatively) whilst 5L had 7% lower ceiling effects when compared to 3L.

5.2.3.1. EQ-5D index, EQ-VAS and combined EQ-5D/EQ-VAS

The differences between the pre-operative and one-year post-operative EQ-5D index and EQ-VAS were calculated and assigned a binary outcome "improved" and "not improved" based on set MCID thresholds. Additionally, the overall combined improvement of the EQ-5D index and EQ-VAS was assessed. Table 5-3 describes the study outcomes.

5.2.3.2. Minimal Clinically Important Difference (MCID)

MCID refers to the smallest improvement deemed significant and meaningful by a patient and vary by patient cohort, conditions, and instrument (403, 404). Several methods exist to calculate this, including: paired-data specific; anchor-based; and distribution-based MCIDs. The choice of which value to select depends on the nature of the study and the research question. Previous studies selected MCID for the EQ-5D index using either the Walters *et al.'s* (405) anchor-based MCID of 0.074, or the statistical effect size of Cohen's d MCID of 0.2 (406). However, these thresholds are uncertain and not specific to hip osteoarthritis or THR treatment (397).

More recently, Kang measured the internal responsiveness of the EQ-5D index for hip replacement using NHS PROMs data between 2009 and 2015 (n = 181,424) and a variety of methods including univariate and multivariate standardised response mean (SRM), standardised effect size (SES) and paired data-specific MCID whilst adjusting for baseline co-variates such as age, gender, and co-morbidities (407). They reported EQ-5D index MCID values of **0.196** for the SRM-applied method with a desired effect size equivalent to Cohen's medium effect (0.5) - a value exceeding 0.196 signifies a meaningful improvement from a clinical perspective. The adjusted multivariate capacity of benefit score, using the paired-data method with applied SRM as a desired effect size, was measured to be **0.428**. The latter represents the magnitude of improvement that would be considered of substantial clinical benefit.

In terms of EQ-VAS, a recent study applied machine learning models and derived an MCID value of **7.81** using the change difference method in a sample of 1,843 total hip arthroplasties with an AUC of 0.84 (399), which is within the range of a Danish study, which used multiple anchor-based approaches and reported a calculated hip minimal MCID range between 5 and 23 (408). Table 5-2 highlights the designated MCIDs in this study.

Outcome	MCID
EQ-5D, SRM	0.196
EQ-5D, Capacity of benefit	0.428
EQ-VAS	7.81

Table 5-2 MCID thresholds per outcome

5.2.4. Model development, comparison, and internal validation

Six machine learning algorithms were developed, namely random forest (RF), gradient boosting machine (GBM), penalized logistic regression (with Lasso and Ridge penalty), and classification tree (with and without pruning), and then compared to traditional statistical model logistic regression to predict the improvement in one-year PROMs. The SHAR derivation cohort of 82,526 observations was used for model development and internal validation. The data extract was randomly split into a training dataset (N=66,020, 80% of the patients) and an internal validation cohort (N=16,506, 20% of the patients). The training cohort was used to train the ML models and to adjust their hyperparameters via cross-validation whereas the validation cohort was used to assess the models' performance on unseen data. Model performance was evaluated and compared in terms of area under the receiver operating characteristic curve (AUC) in each dataset. The top performing machine learning model was used to identify the key features driving its performance.

5.2.5. Statistical analysis and feature importance

Descriptive statistics including mean, median, and percentage were used to describe the rates of one-year PROMs changes, and across various prosthesis groups. R statistical computing environment version 4.3.0 R packages was used for data analyses. The packages used for machine learning model development are mentioned in Chapter 3 (Section 3.2.5).

5.3. Results

5.3.1. Summary statistics of outcomes

Descriptive figures related to the changes in one-year PROMs across the different thresholds are presented in Table 5-3. Using the SRM cut-off, approximately two-thirds of patients reported a one-year improvement in each HRQoL domain (EQ-5D index = 66.3%, EQ-VAS = 69.1%). This decreased to 51.7% when we combined improvements between the two outcomes. Using a higher cut-off value for EQ-5D index yielded lower rates (~40% for the EQ-5D index and 31.3 for the combined one).

Outcome at one-year	Derivation cohort (N=82,526)	
	N	%
EQ-5D index, SRM threshold (Improved)	54,734	66.3
EQ-5D index, CoB threshold (Improved)	32,890	39.9
EQ-VAS (Improved)	57,064	69.1
Combined SRM EQ-5D^{SRM} and EQ-VAS (Improved)	42,653	51.7
Combined EQ-5D^{CoB} and EQ-VAS (Improved)	25,849	31.3

Table 5-3 SHAR study outcomes

Table 5-4 shows the percentages of improvements per prosthesis group across the five outcomes. Overall, hip resurfacing or uncemented and reverse hybrid fixations were variably observed to have the highest rates of improvements in one-year PROMs.

Prosthesis group	EQ-5D^{CoB} improvement n (%)	EQ-5D^{SRM} improvement n (%)	EQ-VAS Improvement n (%)	Combined EQ-5D^{CoB} + EQVAS n (%)	Combined EQ-5D^{SRM} + EQVAS n (%)
Cemented	21,289 (39.2)	34,750 (64.0)	36,834 (67.8)	16,440 (30.2)	26,789 (49.3)
Uncemented	6,088 (40.4)	10,777 (71.6)	10,807 (71.8)	4,956 (32.9)	8,574 (56.9)
Hybrid	850 (45.3)	1,335 (71.2)	1,213 (64.7)	630 (33.6)	979 (52.2)
Reverse hybrid	4,442 (41.9)	7,326 (69.2)	7,631 (72.0)	3,623 (34.2)	5,848 (55.2)
Resurfacing	221 (29.0)	546 (71.6)	579 (75.9)	200 (26.2)	463 (60.7)

Table 5-4 Outcomes per prosthesis group

5.3.2. ML models training and internal validation

Prediction performance of the machine learning algorithms in the internal validation cohort are presented in Table 5-5. Gradient boosting machine (GBM) was consistently the best performing machine learning algorithm in predicting significant clinical improvements in one-year PROMs (AUCs range 0.80 – 0.97%), outperforming the other models by a narrow margin.

Outcome	Model performance: AUC % (95% CI)						
	Random forest	Gradient boosting machine	Ridge regression	Lasso regression	Logistic regression	Classification tree	Classification tree with pruning
EQ5D index, SRM	0.82 (0.81-0.83)	**0.83 (0.82-0.84)**	0.81 (0.80-0.82)				0.79 (0.78-0.79)
EQ5D index, CoB	0.96 (0.95-0.96)	**0.97 (0.96-0.97)**	0.94 (0.94-0.95)	0.96 (0.95-0.96)	0.95 (0.95-0.96)	0.96 (0.96-0.96)	0.96 (0.95-0.96)
EQVAS	0.80 (0.79-0.81)	**0.82 (0.81-0.82)**	0.80 (0.80-0.81)	0.80 (0.80-0.81)	0.80 (0.80-0.81)	0.79 (0.78-0.80)	0.75 (0.74-0.76)
Combined, CoB	0.94 (0.94-0.95)	**0.95 (0.95-0.95)**	0.93 (0.93-0.93)	0.93 (0.93-0.93)	0.93 (0.93-0.94)	0.94 (0.94-0.94)	0.93 (0.93-0.94)
Combined, SRM	0.79 (0.78-0.80)	**0.80 (0.79-0.81)**	0.76 (0.75-0.76)	0.76 (0.75-0.77)	0.76 (0.75-0.77)	0.78 (0.77-0.78)	0.70 (0.69-0.71)

Table 5-5 Prediction performance of all applied machine learning algorithms in the internal validation cohort. *The best-performing models for each outcome are highlighted in bold font*
AUC = area under the receiver operating characteristic curve, CI = confidence interval

5.3.3. Feature importance and response analysis

As the GBM was the top-performing model, we proceeded to examine the responses and the significance of the features identified by this model. Whilst these features provide insights into the model internal decision-making process, caution is strongly advised

when interpreting the results since feature importance is not subject to inferential testing and does not establish statistical significance (see Section 5.4.2).

Three features consistently emerged as key drivers for the predictions of changes in the EQ-5D index (SRM and CoB thresholds), and EQ-VAS, and these include in ranked sequence: pre-operative EQ-5D index (importance: SRM 73.1%; and 95.8% for CoB) and EQ-VAS (importance 72.2%), respectively; pre-operative Charnley classification (importance: SRM 9.0%; CoB 0.9%; and 6.3% for EQ-VAS) and age (importance: SRM 3.3%; CoB 0.6%; and 4.2% for EQ-VAS) (Figure 5-1).

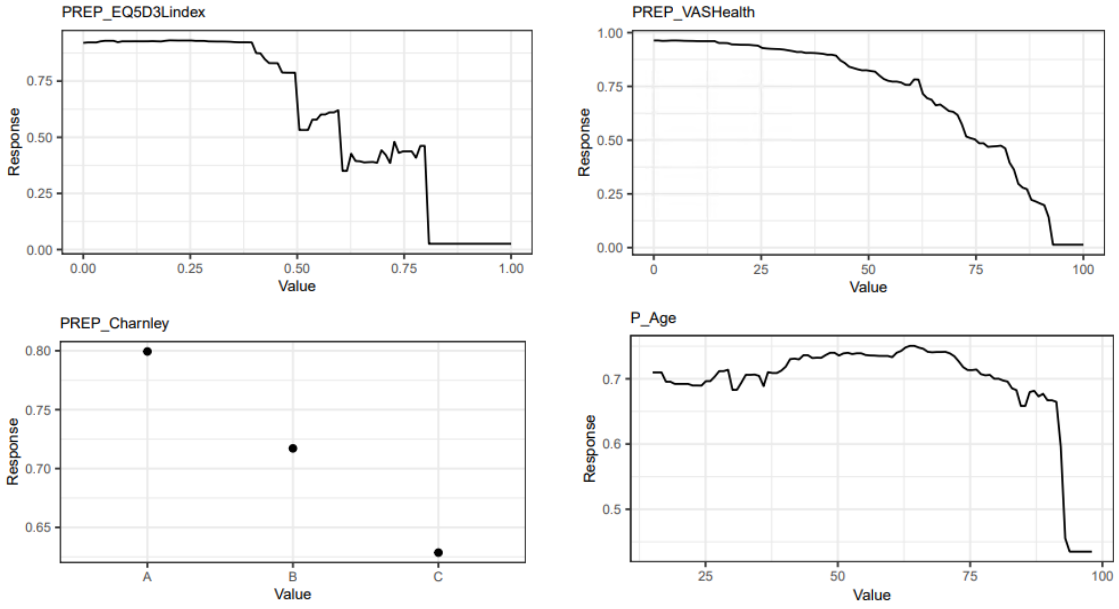
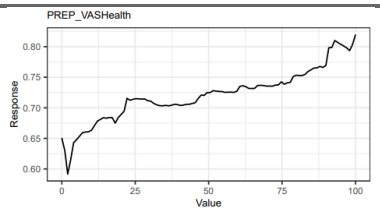
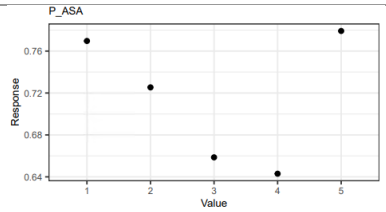
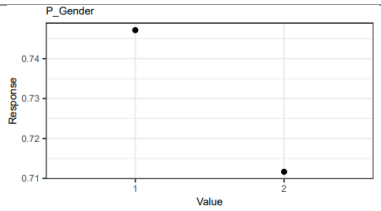
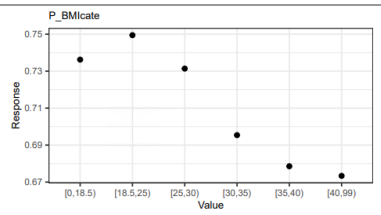
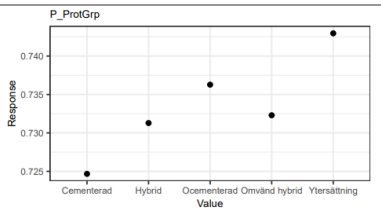
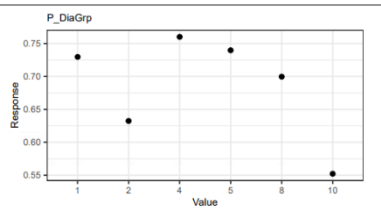


Figure 5-1 The three most important features for the GBM predictive models
The top two are the number one feature for EQ-5D index and EQ-VAS outcomes, respectively. Pre-operatively Charnley classification and age were the second and third ranked features, respectively

A lower pre-operative PROMs score appeared to be linked with better post-operative improvements. On the other hand, age below 75 years tended to have a similar improvement in HRQoL but this sharply declined with increasing age. Given the high importance of these three features in driving the EQ-5D CoB threshold model (totalling 97.3%), we excluded further feature analysis for the remaining 2.7%. Table 5-6 identified the remaining six key features driving the predictions for changes in EQ-5D index (SRM threshold) and EQ-VAS.

EQ-5D index, SRM

Rank	Feature importance (%)	Response graph
4	Pre-operative VAS score (3.2%)	
5	ASA (2.6%)	
6	Gender (1.36%)	
7	BMI (1.34%)	
8	Prosthesis group (0.8%)	
9	Diagnosis group (0.7%)	

EQ-VAS

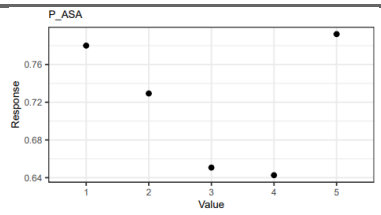
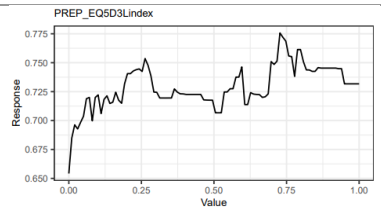
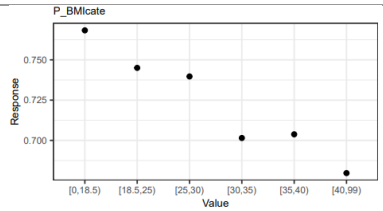
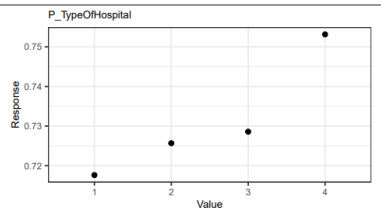
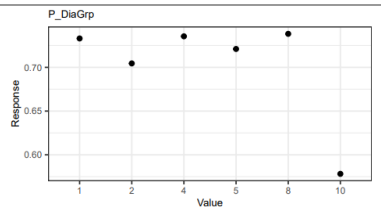
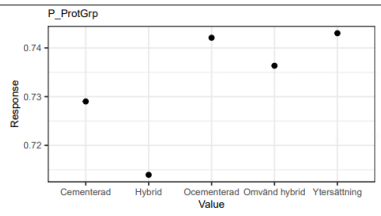
Rank	Feature importance (%)	Response graph
4	ASA (3.7%)	
5	Pre-operative EQ-5D index score (3.3%)	
6	BMI (1.4%)	
7	Unit type (0.7%)	
8	Diagnosis group (0.6%)	
9	Prosthesis group (0.4%)	

Table 5-6 Feature importance for EQ-5D index (left) and EQ-VAS (right) GBM models

There appeared to be a positive correlation between two arms of the EQ-5D (index and VAS), wherein higher pre-operative values in one arm seemed to improve one-year PROMs in the other. In both outcomes, the extremes of ASA grades were identified by the model to augment one-year PROMs, whilst obesity classes I – III were observed to have

less pronounced improvements. Hip resurfacing and uncemented and reverse hybrid fixation were shown by the model to have the highest response effect on one-year PROMs. Care delivered at private units was identified to have better improvements in one-year EQ-VAS scores but not E5-5D. There were similar improvement levels among all surgical indications studied, except for the "other" group.

5.4. Discussion

5.4.1. Summary of results

This study has shown that predicting one-year improvements in HRQoL following primary elective hip arthroplasty using machine learning algorithms achieved a good to excellent binary discriminative power and slightly outperformed conventional statistical predictive models. This is the first study to incorporate clinical and implant-specific registry data to generate personalised stratification models, which if applied in clinical practice, may improve the shared decision-making process.

The descriptive statistics alludes to the fact that approximately two-thirds of individuals reach a significant clinical improvement at one-year post-operatively in the health utility index (66.3%) and general health (69.1%), whilst only 40% reached substantial meaningful improvement in their EQ-5D index score at one-year. Our figures are comparable to previous studies (398, 399, 409), reporting 53.4%[409] - 58%[399] and 44.5%[398] - 66.3%[399] of patients reaching EQ-5D and EQ-VAS MCIDs, respectively, using similar cut-off levels. This denotes the population-level consistency and generalisability of our dataset.

With regards to EQ-VAS, our machine learning models showed strong classification performance in identifying patients with meaningful clinical improvement following THR after one year, equating 0.82 AUC for EQ-VAS using the GBM model. The model found

that lower pre-operative VAS health and Charnley categories and ages less than 75 years to be predictor factors of general health improvement, totalling around 83% of feature importance. Huber *et al.* used the NHS PROMs hip dataset (n = 62,429) between 2015 and 2017 to predict changes in EQ-VAS with a minimal important difference (MID) threshold of 11 points (398). The authors reported an AUC value of 0.87 for the Gradient Boosting Machine model outperforming linear models and highlighted that pre-operative VAS scores and limping to be the key predictors behind their algorithms. While it aligns with our findings, the lack of clinical and implant-specific data and the use of half a standard deviation of baseline preoperative-VAS score as a MID threshold, rather than the standardised response mean or paired data-specific MCID, may have limited the clinical applicability and generalisability of their models.

In terms of health utility (EQ-5D) index, the GBM models reached AUC values of 0.83 and 0.97 in forecasting significant and substantial clinical improvements at one year. Similar to EQ-VAS, this predictive performance heavily relied on the three aforementioned features, with the caveat of pre-operative EQ-5D index score rather than EQ-VAS, reaching 85.4% - 97.3% feature importance for significant and substantial improvements, respectively. This agrees with a recent German study, which provided instrumental calculations for the MCID of 7.83 for EQ-VAS which was implemented in this study (399). The research team analysed 1,843 THR observations collected from nine hospitals in Germany with AUC values of 0.81 for the EQ-5D index and 0.84 for the EQ-VAS, echoing prior results. However, the authors acknowledged the need for a larger and representative sample size to enhance the models' external validity.

When we predicted overall combined improvement of EQ-5D and EQ-VAS, our predictive accuracy marginally declined by approximately 2%, permitting further differentiation of patients likely to have higher HRQoL improvements postoperatively. While the majority

of GBM's predictive ability stemmed from these three features, our models highlighted additional factors possibly contributing to the predictions. Of particular importance, obese patients were noted to exhibit lower improvement scores, while procedures such as hip resurfacing and uncemented femoral fixation displayed the highest response effect. The latter observations were also evident in our descriptive figures. Interpreting these findings in light of prior literature (see Sections 1.9 and 1.11), previous studies tend to learn towards excluding BMI as a predictor of functional improvements. However, it is important to note that most of these studies focused on condition-specific PROMs rather than HRQoL. In contrast, while some authors have attributed the increased safety portfolio of HRA to patient selection bias (12, 14), there is strong evidence to support superior functional outcomes in HRA (see Section 1.9).

5.4.2. Limitations and further work

Whilst the limitations related to the retrospective observational design using a single national dataset remain, the alignment with prior research findings, in terms of descriptive statistics and AUC values, may indicate a degree of population-level consistency. The lack of control groups prevented the construct of patient trajectories without THR, meaning the study cannot discern whether the limited improvements post-operatively were solely attributed to the procedure itself or if THR prevented a substantial deterioration in HRQoL. Moreover, it is important to acknowledge that the PROMs only offer a snapshot assessment of health-related quality of life on a single day, potentially not accounting for temporal variations. Similarly, while standardised cut-off MCID levels were carefully employed, different thresholds may yield different results, whereas the choice of a binary outcome (improvement or no improvement) may have restricted the models' ability in forecasting the extent of changes. Finally, although the features identified by the best performing model provided insights into its ensemble internal decision-making process,

this approach is not subject to inferential testing assumptions and therefore inherently limited in evaluating associations. Future studies should seek to externally validate our algorithms on other datasets, such as the NJR for England and Wales, as well as forecast improvements using condition-specific PROMs such as the OHS or HHS, which are not captured by the SHAR. Prospective clinical trials are needed to assess the applicability and accuracy of these predictions using a mix of PROMs in clinical settings.

5.4.3. Conclusions

This is the first study to incorporate patient, surgical and implant-specific factors to forecast one-year improvements in HRQoL measures following primary elective THR. The internally validated machine learning models for each outcome marginally outperformed the logistic regression model and showed good to excellent discriminative power in predicting significant and substantial clinical improvements, respectively. Three features appeared to account for over 80% of the algorithmic power in the ensemble gradient boosting machine (GBM) model including: pre-operative baseline outcome score; pre-operative Charnley classification; and age. Additional research is needed to investigate the optimal set of PROMs, such as EQ-5D, OHS and HHS, that aid surgical decision-making.

Chapter 6 Predicting revision risk due to periprosthetic fracture

6.1. Introduction

Chapter 4 of this thesis highlighted the key determinants for short- and long-term revision risks following primary hip arthroplasty. This includes, besides patient and diagnostics factors, implant selection. Cementless femoral fixation was significantly associated with a higher revision risk at two years postoperatively. Previous studies reported a higher incidence of periprosthetic fractures (PPF) after press-fit techniques (410-412) with intraoperative PPFs fourteen times more likely to occur postoperatively with cementless implants (410).

Examination of the SHAR's 2022 annual report reveals that periprosthetic fracture (PPF) is the third leading cause for first-time revision hip arthroplasty (13.2%) following aseptic loosening (44%) and infection (24.5%). The report lacks specific information regarding the periprosthetic fracture risk comparison between cemented and uncemented fixation in primary elective hip arthroplasty. However, it indicates that the risk of implant fracture, which is a rare complication involving a fracture of the prosthesis stem, is increased in the use of certain cemented stems in comparison to cementless fixation (0.06% versus 0.02%) (5). In contrast, a previous study conducted by the Swedish Hip Arthroplasty Register between 1992 and 2007 found a higher risk of periprosthetic stem fracture following primary total hip arthroplasty in cementless fixation compared to cemented fixation (17% versus 6%; Adjusted RR 8.0 [4.5–14]) (345).

The annual report of the National Joint Registry of England and Wales in 2022 yields similar findings with PPF being the four predominant causes for revision being aseptic loosening (24.6%), dislocation/subluxation (17.4%), periprosthetic fracture (15.7%) and

infection (15.2%). The annual report shows a higher incidence of revisions per 1,000 prosthesis-years for periprosthetic fracture following primary hip arthroplasty in cementless fixation (0.67%) than cemented fixation (0.53%) with hybrid fixation (0.86%) and reverse hybrid fixation (0.66%) also displaying relatively elevated rates (6).

According to the Australian Orthopaedic Association National Joint Replacement Registry's (AOANJRR) annual report of 2022 (7), PPF is the third leading reason for revision following primary total hip arthroplasty (21.8%), after infection (22.7%) and prosthesis dislocation/instability (22.0%). AOANJRR reports no difference in revision rate between cemented and hybrid fixations for all age groups, whilst cementless fixation has a higher revision risk following primary THR compared to hybrid and cemented fixation in patients aged 75 and above, and a time-dependant increased revision risk for the 55-64 (first month postoperatively) and 65-74 (1.5 years postoperatively) age groups.

As outlined in Chapter 3, machine learning models exhibit a superior predictive ability to better forecast the occurrence of events than traditional statistics and offer a new perspective taking into consideration varied patient and surgical factors and their complex interactions. Although relatively rare, the occurrence of postoperative periprosthetic fractures is on the rise and presents significant challenges in the context of primary hip arthroplasty. These complications are linked to a heightened risk of negative health outcomes, increased mortality rates, significant financial burdens, and pose technical difficulties for surgeons (413, 414).

The aim of this chapter is to develop ML-based risk prediction models and compare their predictive ability to traditional statistical linear models using large datasets from the Swedish Hip Arthroplasty Register. The outcome measure will be to predict the risk of re-operation and revision due to periprosthetic fracture at multiple time points ranging from

thirty days to one year postoperatively using the SHAR dataset. Chapter 7 employs a propensity score matching technique to investigate the effect of implant fixation on periprosthetic fracture outcome.

6.2. Patients and Methods

6.2.1. Data sources

This study uses the SHAR datasets described previously in Chapter 2. All patients who underwent primary elective hip arthroplasty in the SHAR extract between 2008 and 2018 (n=154,539) were included. Twenty cases related to acetabular periprosthetic fractures were excluded (n=154,519). This study is in two parts; the first part uses the SHAR dataset to develop and internally validate the machine learning algorithms in predicting revision and re-operation outcomes due to periprosthetic fracture, followed by a comparison of the key features driving the predictions for the top performing machine learning model and the traditional statistical model, logistic regression. Table 6-1 provides the variables used in the SHAR dataset.

		SHAR
Patient factors	Age	
	Gender	
	BMI	
	ASA grade	
	Side of operation	
	Hospital type	
Surgical factors	Operation type	
	Diagnosis group	
	Incision type	
	Prothesis group	
	Cement type	
Cup implant factors	Fixation	
	Resurfacing	
	Articulation	
	Modular or Monoblock	
	Inner diameter	
	Modification	
Femoral implant factors	Fixation	
	Stem resurfacing	
	Articulation	
	Head size	
Outcomes	Revision due to periprosthetic fracture	30-day
		60-day
		90-day
		1-year
	Re-operation due to periprosthetic fracture	30-day
		60-day
		90-day
		1-year

Table 6-1 Sources of data

6.2.2. Exposures

For both parts of the study, machine learning models were applied to forecast the outcomes of hip arthroplasty. Analyses were restricted to patients undergoing elective primary hip replacement. Patients who had undergone hip arthroplasty due to specific diagnoses such as trauma, tumour-related conditions, and degenerative conditions other than osteoarthritis were excluded. Patients below the age of 18 were excluded to minimise

potential confounding factors that could arise from the different outcomes observed in paediatric patients (369-371).

6.2.3. Outcome of interest

The primary outcome measure in the SHAR dataset was femoral periprosthetic fracture risk leading to either a revision or re-operation procedure within 30, 60, 90 days and 1 year postoperatively. As described before, re-operation is defined as any additional surgery to the hip, regardless of the actions taken to any of the implant components (replaced, extracted, or left untouched) whereas revision is defined as a re-operation where at least one component is exchanged, extracted from, or added to the prosthesis. Several machine learning models (described below in Section 6.2.4) were developed and compared based on their accuracy in predicting revision risk and re-operation due to periprosthetic fracture following elective primary hip arthroplasty at 30, 60, 90 days and 1-year intervals (Table 6-1).

6.2.4. Model development, comparison, and internal validation

We applied six machine learning algorithms, namely random forest (RF), gradient boosting machine (GBM), penalised logistic regression (with Lasso and Ridge penalty), and classification tree (with and without pruning), and compared them to traditional statistical model logistic regression to predict the occurrence of intra-/post-operative periprosthetic fracture necessitating a revision procedure. The SHAR derivation cohort of 154,519 observations was used for model development and internal validation. The data extract was randomly split into a training dataset (N=123,615, 80% of the patients) and an internal validation cohort (N=30,904, 20% of the patients). The training cohort was used to train the ML models and to adjust their hyperparameters via cross-validation whereas the validation cohort was used to assess the models' performance on unseen data. For a more comprehensive understanding of each variable and a detailed summary of the

SHAR dataset, please refer to Appendix 2. Model performance was evaluated and compared in terms of area under the receiver operating characteristic curve (AUC) in each dataset. The top performing machine learning model was used to identify the key features driving its performance.

6.2.5. Statistical analysis and feature importance

Descriptive statistics, including mean and percentage, were used to describe the rates of implant fracture across various age and sex groups as well as per year in the SHAR datasets. Kaplan-Meier (KM) non-parametric estimates (372) were employed to describe the rates of periprosthetic fracture leading to revision or re-operation procedures across various age and sex groups in the SHAR dataset. As before, analyses were performed using R statistical computing environment version 4.3.0. R packages used for the machine learning model development are mentioned in Chapter 3 (Section 3.2.5).

6.3. Results

6.3.1. Summary statistics of outcomes

Descriptive figures related to the SHAR femoral periprosthetic fracture outcome are presented in Table 6-1. The one-year rates for revision and re-operation were 0.17% and 0.23%, respectively.

Outcome	Time from primary operation	Derivation cohort (N=154,519) N	%
Revision due to periprosthetic fractures	30 days	120	0.07
	60 days	169	0.10
	90 days	196	0.12
	1 year	266	0.17
Re-operation due to periprosthetic fractures	30 days	139	0.08
	60 days	202	0.13
	90 days	239	0.15
	1 year	369	0.23

Table 6-2 SHAR study outcomes

Table 6-3 provides Kaplan-Meier estimates with 95% confidence intervals for the percentage of patients experiencing revision due to periprosthetic fractures at different time intervals following primary elective hip arthroplasty, stratified by gender and age groups. Overall, one-year revision rates due to fracture ranged from 0.00% to 0.23% with no apparent discrepancy between gender groups.

			30-day revision		60-day revision		90-day revision		1-year revision	
		N total	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)
All patients		154,519	120	0.05% (0.02-0.10)	169	0.08% (0.03-0.13)	196	0.10% (0.05-0.15)	266	0.13% (0.07-0.19)
Male	All	66,830	50	0.05% (0.02-0.10)	68	0.08% (0.03-0.13)	84	0.10% (0.05-0.16)	130	0.15% (0.08-0.22)
	18-34 years	393	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)
	35-49 years	4,244	1	0.02% (0.00-0.07)	2	0.05% (0.00-0.11)	3	0.07% (0.00-0.15)	5	0.12% (0.02-0.23)
	50-59 years	11,264	12	0.11% (0.05-0.17)	18	0.16% (0.09-0.24)	21	0.19% (0.11-0.27)	24	0.22% (0.13-0.31)
	60-69 years	22,783	15	0.07% (0.03-0.10)	21	0.09% (0.05-0.13)	26	0.12% (0.07-0.16)	49	0.23% (0.16-0.29)
	70-79 years	21,276	17	0.08% (0.04-0.12)	19	0.09% (0.05-0.13)	24	0.11% (0.07-0.16)	41	0.20% (0.14-0.26)
	≥ 80 years	6,870	5	0.07% (0.01-0.14)	8	0.12% (0.04-0.20)	10	0.15% (0.06-0.24)	11	0.17% (0.07-0.26)
Female	All	87,689	70	0.06% (0.02-0.10)	101	0.09% (0.04-0.14)	112	0.10% (0.05-0.15)	136	0.12% (0.06-0.17)
	18-34 years	415	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)
	35-49 years	3,325	2	0.06% (0.00-0.14)	3	0.09% (0.00-0.19)	3	0.09% (0.00-0.19)	3	0.09% (0.00-0.19)
	50-59 years	10,903	12	0.11% (0.05-0.17)	19	0.18% (0.10-0.26)	20	0.19% (0.10-0.27)	21	0.20% (0.11-0.28)
	60-69 years	27,548	32	0.12% (0.08-0.16)	40	0.15% (0.10-0.19)	42	0.15% (0.11-0.20)	52	0.19% (0.14-0.24)
	70-79 years	32,267	21	0.07% (0.04-0.09)	33	0.10% (0.07-0.14)	39	0.12% (0.08-0.16)	46	0.15% (0.10-0.19)
	≥ 80 years	13,231	3	0.02% (0.00-0.05)	6	0.05% (0.01-0.08)	8	0.06% (0.02-0.10)	14	0.11% (0.05-0.17)

Table 6-3 Summary and Kaplan-Meier (KM) estimates for the cumulative incidence of revision outcomes due to periprosthetic fractures at each time point by age and sex using the SHAR

Table 6-4 highlights the temporal variations in Kaplan-Meier estimates for revision due to periprosthetic fracture over successive years between 2008 and 2018. No statistically significant differences were detected throughout all the years ($p > 0.05$).

		30-day revision			60-day revision			90-day revision			1-year revision		
Year	N total	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§
2008	11,058	6	0.05%	1.00	10	0.09%	1.00	10	0.09%	1.00	15	0.14%	1.00
2009	13,026	16	0.12%	2.26 (0.88-5.79)	18	0.14%	1.53 (0.70-3.31)	20	0.15%	1.69 (0.79-3.62)	32	0.25%	1.80 (0.97-3.33)
2010	13,547	4	0.03%	0.54 (0.15-1.93)	8	0.06%	0.65 (0.25-1.66)	10	0.07%	0.81 (0.34-1.96)	17	0.13%	0.92 (0.46-1.85)
2011	13,699	10	0.07%	1.33 (0.48-3.67)	15	0.11%	1.20 (0.54-2.68)	19	0.14%	1.52 (0.71-3.28)	21	0.15%	1.12 (0.58-2.18)
2012	13,946	11	0.08%	1.45 (0.53-3.94)	11	0.08%	0.87 (0.37-2.06)	14	0.10%	1.11 (0.49-2.50)	22	0.16%	1.16 (0.60-2.23)
2013	13,987	15	0.11%	1.98 (0.77-5.12)	19	0.14%	1.51 (0.70-3.25)	22	0.16%	1.74 (0.82-3.69)	31	0.22%	1.63 (0.88-3.03)
2014	14,226	17	0.12%	2.22 (0.87-5.65)	23	0.16%	1.80 (0.86-3.79)	23	0.16%	1.80 (0.85-3.78)	31	0.22%	1.61 (0.87-2.98)
2015	14,410	12	0.08%	1.54 (0.58-4.12)	22	0.15%	1.70 (0.80-3.56)	24	0.17%	1.84 (0.88-3.86)	27	0.19%	1.38 (0.73-2.60)
2016	15,000	8	0.05%	0.99 (0.34-2.85)	11	0.07%	0.81 (0.34-1.92)	16	0.11%	1.18 (0.53-2.60)	25	0.17%	1.23 (0.65-2.34)
2017	15,515	9	0.06%	1.08 (0.38-3.03)	10	0.06%	0.71 (0.29-1.72)	14	0.09%	1.00 (0.44-2.25)	19	0.12%	0.91 (0.46-1.79)
2018	16,105	12	0.08%	1.42 (0.53-3.78)	22	0.15%	1.60 (0.75-3.38)	24	0.16%	1.78 (0.85-3.74)	26	0.19%	1.53 (0.81-2.89)

Table 6-4 Change in revision rates due to periprosthetic fracture by year of primary operation in the SHAR
§ Adjusted for Age and Gender
KM = Kaplan-Meier; HR = Hazard Ratio; CI = confidence interval
No statistically significant differences were identified across years

Kaplan-Meier estimates, stratified by age and gender, for the cumulative incidence of re-operation due to periprosthetic fractures at four intervals are presented in Table 6-5. No observed differences were found between males and females, whilst the incidence of fractures necessitating re-operation was highest in the 50-69 age groups.

			30-day re-operation		60-day re-operation		90-day re-operation		1-year re-operation	
		N total	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)	N	KM estimate (95% CI)
All patients		154,519	139	0.07% (0.02-0.11)	202	0.10 (0.05-0.16)	239	0.12% (0.06-0.19)	369	0.20% (0.12-0.28)
Male	All	66,830	53	0.06% (0.02-0.10)	76	0.09% (0.04-0.14)	97	0.12% (0.06-0.19)	166	0.21% (0.12-0.29)
	18-34 years	393	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)
	35-49 years	4,244	1	0.02% (0.00-0.07)	2	0.05% (0.00-0.11)	4	0.10% (0.00-0.19)	8	0.20% (0.06-0.33)
	50-59 years	11,264	13	0.12% (0.05-0.18)	21	0.19% (0.11-0.27)	25	0.23% (0.14-0.32)	28	0.26% (0.16-0.35)
	60-69 years	22,783	16	0.07% (0.04-0.11)	23	0.10% (0.06-0.14)	28	0.13% (0.08-0.17)	59	0.27% (0.20-0.34)
	70-79 years	21,276	18	0.09% (0.05-0.12)	21	0.10% (0.06-0.14)	28	0.13% (0.08-0.18)	54	0.27% (0.20-0.34)
	≥ 80 years	6,870	5	0.07% (0.01-0.14)	9	0.13% (0.05-0.22)	12	0.18% (0.08-0.28)	17	0.26% (0.14-0.38)
Female	All	87,689	86	0.08% (0.03-0.12)	126	0.12% (0.06-0.18)	142	0.13% (0.07-0.19)	203	0.20% (0.12-0.28)
	18-34 years	415	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)	0	0.00% (0.00-0.00)
	35-49 years	3,325	3	0.09% (0.00-0.19)	5	0.15% (0.02-0.28)	5	0.15% (0.02-0.28)	9	0.28% (0.10-0.46)
	50-59 years	10,903	13	0.12% (0.05-0.18)	20	0.19% (0.10-0.27)	21	0.19% (0.11-0.28)	23	0.21% (0.13-0.30)
	60-69 years	27,548	37	0.13% (0.09-0.18)	49	0.18% (0.13-0.23)	52	0.19% (0.14-0.24)	71	0.26% (0.20-0.33)
	70-79 years	32,267	24	0.07% (0.04-0.10)	39	0.12% (0.08-0.16)	46	0.14% (0.10-0.19)	69	0.22% (0.17-0.27)
	≥ 80 years	13,231	9	0.07% (0.02-0.11)	13	0.10% (0.05-0.15)	18	0.14% (0.07-0.20)	31	0.24% (0.16-0.33)

Table 6-5 Summary and Kaplan-Meier (KM) estimates for the cumulative incidence of re-operation outcomes due to periprosthetic fractures at each time point by age and sex using the SHAR

Table 6-6 illustrates the fluctuations in Kaplan-Meier estimates for re-operation due to periprosthetic fracture between 2008 and 2018. No statistically significant differences were detected throughout all the years ($p > 0.05$).

		30-day re-operation			60-day re-operation			90-day re-operation			1-year re-operation		
Year	N total	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§	N	KM estimate	HR (95% CI)§
2008	11,058	9	0.08%	1.00	13	0.12%	1.00	14	0.13%	1.00	27	0.25%	1.00
2009	13,026	19	0.15%	1.79 (0.81-3.97)	24	0.18%	1.57 (0.80-3.08)	26	0.20%	1.57 (0.82-3.02)	48	0.37%	1.50 (0.94-2.41)
2010	13,547	5	0.04%	0.45 (0.15-1.35)	10	0.07%	0.62 (0.27-1.43)	12	0.09%	0.70 (0.32-1.51)	28	0.21%	0.84 (0.49-1.43)
2011	13,699	11	0.08%	0.98 (0.40-2.37)	17	0.12%	1.05 (0.51-2.17)	24	0.18%	1.38 (0.71-2.67)	33	0.24%	0.98 (0.59-1.63)
2012	13,946	12	0.09%	1.06 (0.44-2.52)	12	0.09%	0.73 (0.33-1.61)	15	0.11%	0.85 (0.41-1.76)	27	0.20%	0.79 (0.46-1.35)
2013	13,987	16	0.11%	1.41 (0.62-3.20)	21	0.15%	1.28 (0.64-2.57)	24	0.17%	1.36 (0.70-2.63)	38	0.27%	1.11 (0.67-1.82)
2014	14,226	17	0.12%	1.48 (0.66-3.33)	24	0.17%	1.45 (0.73-2.85)	26	0.18%	1.45 (0.75-2.78)	39	0.28%	1.12 (0.68-1.83)
2015	14,410	13	0.09%	1.12 (0.47-2.62)	26	0.18%	1.54 (0.79-3.01)	28	0.20%	1.54 (0.81-2.93)	37	0.26%	1.05 (0.64-1.73)
2016	15,000	11	0.07%	0.91 (0.37-2.19)	15	0.10%	0.85 (0.40-1.80)	22	0.15%	1.16 (0.59-2.27)	34	0.23%	0.93 (0.56-1.54)
2017	15,515	12	0.08%	0.96 (0.40-2.28)	15	0.10%	0.83 (0.39-1.74)	20	0.13%	1.02 (0.51-2.02)	27	0.18%	0.71 (0.41-1.22)
2018	16,105	14	0.09%	1.10 (0.47-2.55)	25	0.17%	1.40 (0.72-2.75)	28	0.19%	1.50 (0.79-2.85)	31	0.23%	1.06 (0.63-1.78)

Table 6-6 Change in re-operation rates due to periprosthetic fracture by year of primary operation in the SHAR
§ Adjusted for Age and Gender
KM = Kaplan-Meier; HR = Hazard Ratio; CI = confidence interval
No statistically significant differences were identified across years

6.3.2. Machine Learning models training and internal validation

Gradient boosting machine (GBM) was consistently the top performing machine learning algorithm in predicting revision and re-operation risks due to periprosthetic femoral fracture at all time points, except for 30-day revision, in which least absolute shrinkage and selection operator (LASSO) regression outperformed GBM by a narrow margin. Prediction performance of the applied machine learning algorithms in the internal validation cohort are presented in Table 6-7. All models exhibited a remarkably similar area under the curve (AUC) of approximately 0.80.

Outcome	Model performance: AUC in % (95% CI)						
	Random forest	Gradient boosting machine	Ridge regression	Lasso regression	Logistic regression	Classification tree	Classification tree with pruning
Revision due to periprosthetic fracture							
30 days	0.83 (0.79-0.88)	0.84 (0.80-0.88)	0.83 (0.79-0.88)	**0.85 (0.81-0.89)**	0.83 (0.78-0.88)	0.82 (0.78-0.85)	0.82 (0.78-0.85)
60 days	0.81 (0.76-0.86)	**0.86 (0.83-0.89)**	0.85 (0.81-0.89)	0.85 (0.81-0.89)	0.85 (0.81-0.89)	0.84 (0.80-0.87)	0.84 (0.80-0.87)
90 days	0.80 (0.75-0.85)	**0.86 (0.82-0.89)**	0.84 (0.81-0.88)	0.84 (0.80-0.88)	0.82 (0.76-0.87)	0.83 (0.80-0.86)	0.83 (0.80-0.86)
1 year	0.71 (0.65-0.77)	**0.80 (0.75-0.85)**	0.79 (0.74-0.84)	0.79 (0.74-0.84)	0.79 (0.73-0.84)	0.76 (0.71-0.82)	0.76 (0.71-0.82)
Re-operation due to periprosthetic fracture							
30 days	0.77 (0.69-0.84)	**0.84 (0.79-0.89)**	0.83 (0.78-0.88)	0.83 (0.78-0.88)	0.83 (0.77-0.88)	0.80 (0.74-0.85)	0.80 (0.74-0.85)
60 days	0.80 (0.75-0.86)	**0.86 (0.83-0.89)**	0.85 (0.81-0.89)	0.85 (0.81-0.89)	0.85 (0.81-0.89)	0.82 (0.78-0.86)	0.82 (0.78-0.86)
90 days	0.76 (0.69-0.82)	**0.85 (0.81-0.88)**	0.83 (0.80-0.87)	0.83 (0.78-0.87)	0.81 (0.76-0.86)	0.80 (0.76-0.84)	0.80 (0.76-0.84)
1 year	0.66 (0.60-0.73)	**0.79 (0.74-0.84)**	0.78 (0.73-0.83)	0.78 (0.73-0.83)	0.77 (0.72-0.82)	0.71 (0.66-0.76)	0.71 (0.66-0.76)

Table 6-7 Prediction performance of all applied machine learning algorithms in the internal validation cohort. *The best-performing models for each outcome are highlighted in bold font*
AUC = area under the receiver operating characteristic curve, CI = confidence interval

6.3.3. Feature importance and response analysis

The following sub-sections delve into the analysis of response and feature importance for the best-performing machine learning model at each time point. As before, caution is advised when interpreting these findings since they are not subject to inferential testing and statistical significance (see Section 6.4.2).

6.3.3.1. 30-day femoral periprosthetic fracture

Lasso regression and GBM were the best performing ML models for 30-day revision and re-operation due to femoral periprosthetic fracture, respectively. Tables 6-8 and 6-9 identified the key features driving these predictions. The choice of prosthesis, primarily cementless femoral fixation or hip resurfacing, increased fracture risk leading to revision or re-operation whilst patient features, namely obesity class III and higher ASA grades,

played a secondary key part. Age and liped cups were both identified as drivers for increased re-operation risk due to femoral fracture but not revision, which had an increased risk with the use of monoblock stems.

Rank	Feature	Coeff.
1	Cementless femoral fixation	+ 2.93
2	Femoral Monoblock	+ 2.40
3	Resurfacing	+ 1.04
4	Idiopathic necrosis	+ 0.73
5	Obesity class III	+ 0.67
6	ASA grade	+ 0.48
7	Reverse hybrid fixation	+ 0.41

Table 6-8 Feature importance for 30-day revision LASSO model

6.3.2.2 60-day femoral periprosthetic fracture

Increasing age, BMI and ASA variably seemed to increase the incidences of periprosthetic femoral fractures leading to revision or re-operation (Table 6-10). Like the 30-day outcome, cementless femoral fixation, hip resurfacing and reverse hybrid fixation increased fracture risks necessitating an intervention. Large femoral head sizes (> 52 mm) or very small ones (< 25 mm) appeared to be a relatively important feature used in the GBM models, accounting for ~ 5% and ~ 12% of the predictive power for revision and re-operation, respectively.

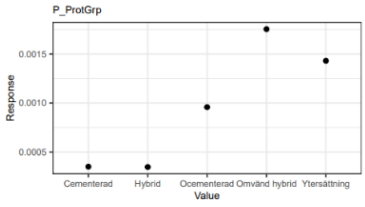
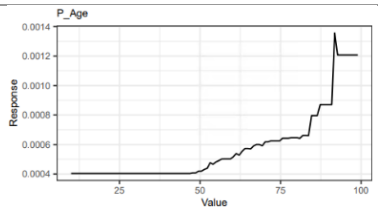
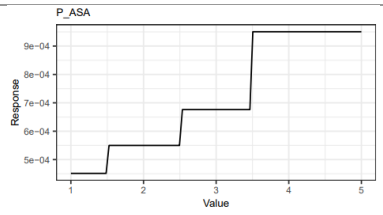
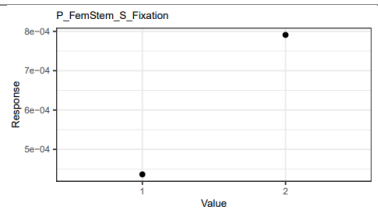
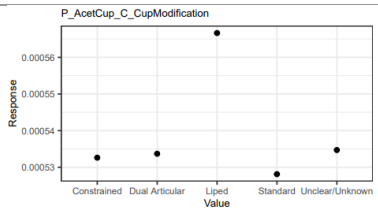
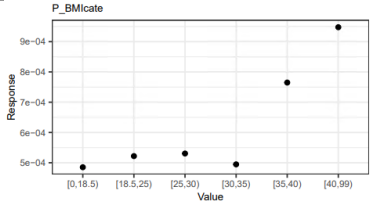
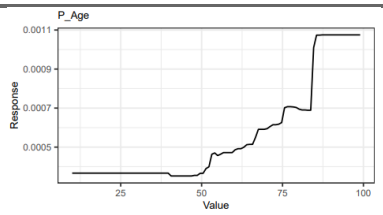
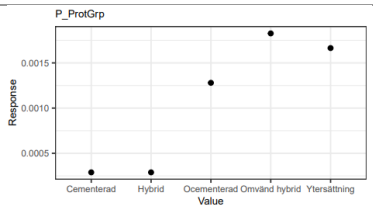
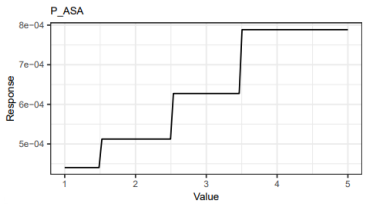
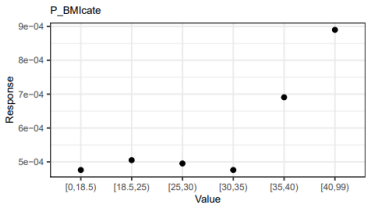
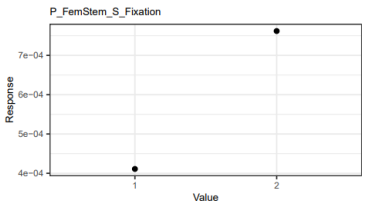
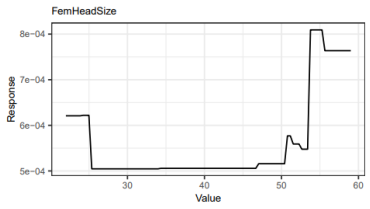
Rank	Feature importance (%)	Response graph
1	Prosthesis group (23%)	
2	Age (21.5%)	
3	ASA (15.8%)	
4	Femoral fixation (6.2 %)	
5	Cup modification (6%)	
6	BMI (5.3%)	

Table 6-9 Feature importance for 30-day re-operation GBM model

Revision

Rank	Feature importance (%)	Response graph
1	Age (43.9%)	
2	Prosthesis group (23.2%)	
3	ASA (8.2%)	
4	BMI (5.2%)	
5	Femoral fixation (4.8%)	
6	Femoral head size (4.8%)	

Re-operation

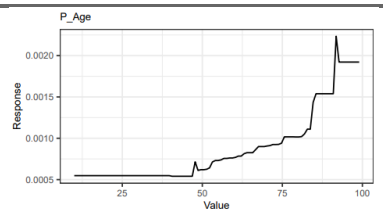
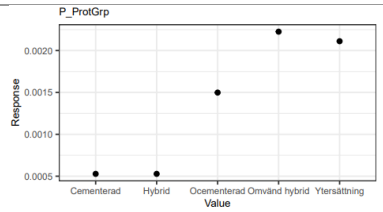
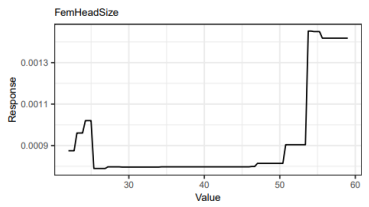
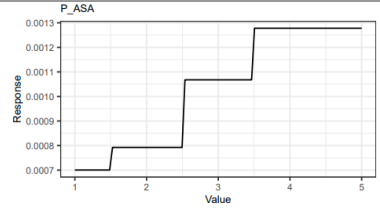
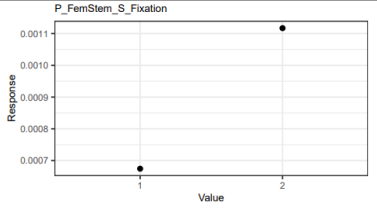
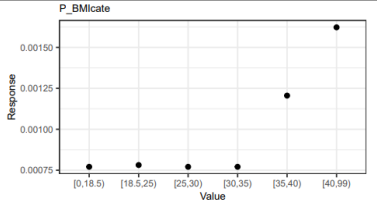
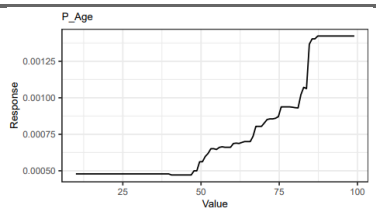
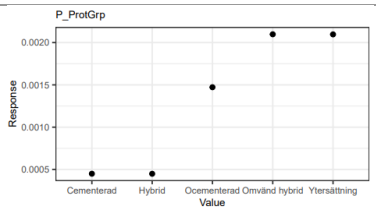
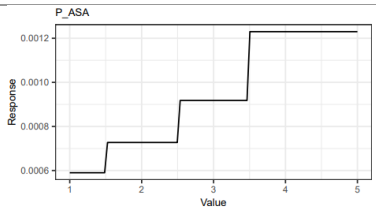
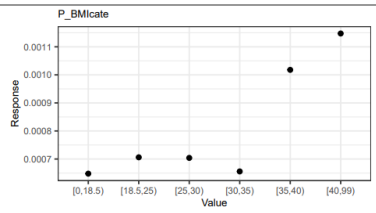
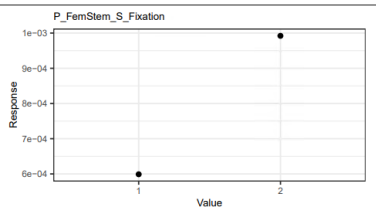
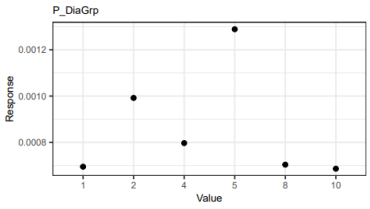
Rank	Feature importance (%)	Response graph
1	Age (29.8%)	
2	Prosthesis group (28.7%)	
3	Femoral head size (11.7%)	
4	ASA (8.3%)	
5	Femoral fixation (4.8%)	
6	BMI (4.5%)	

Table 6-10 Feature importance for 60-day revision (left) and re-operation (right) GBM models

6.3.3.3. 90-day femoral periprosthetic fracture

When compared with the 60-day fracture, no significant changes in response graphs were observed for the 90-day end point (Table 6-11). In the revision model, diagnosis group (idiopathic necrosis), rather than femoral head size, was ranked in the top six features.

Revision

Rank	Feature importance (%)	Response graph
1	Age (36.9%)	
2	Prosthesis group (21.8%)	
3	ASA (10.1%)	
4	BMI (6.6%)	
5	Femoral fixation (5.3%)	
6	Diagnosis group (4.5%)	

Re-operation

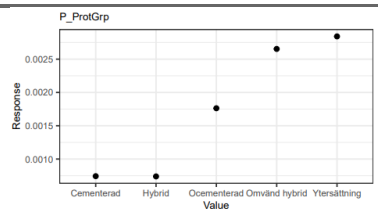
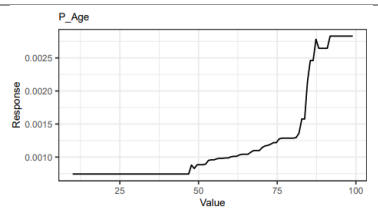
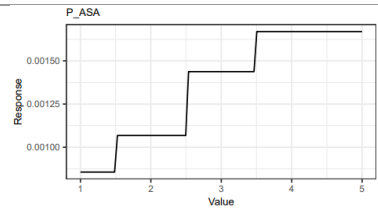
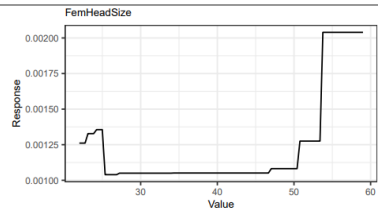
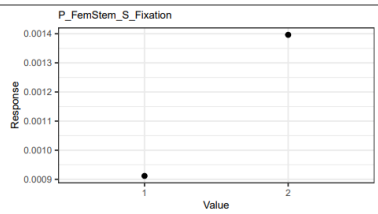
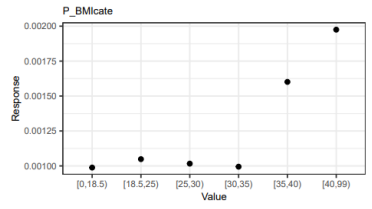
Rank	Feature importance (%)	Response graph
1	Prosthesis group (26.8%)	
2	Age (26.6)	
3	ASA (9.6%)	
4	Femoral head size (9.2%)	
5	Femoral fixation (5.0%)	
6	BMI (4.9%)	

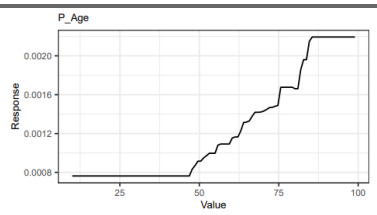
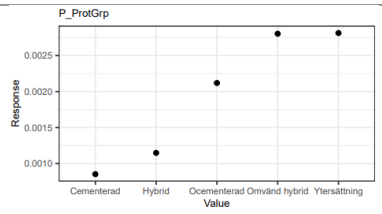
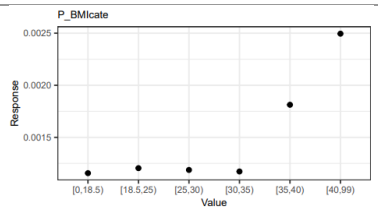
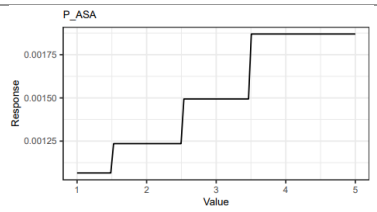
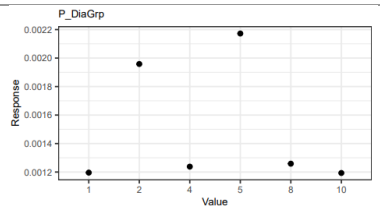
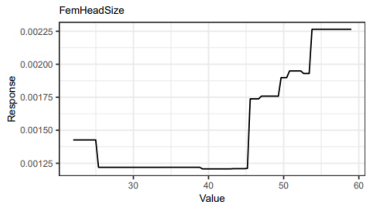
Table 6-11 Feature importance for 90-day revision (left) and re-operation (right) GBM models

6.3.3.4. 1-year femoral periprosthetic fracture

At the one-year interval, pre-operative diagnosis of idiopathic necrosis or inflammatory arthritis became apparent and increased long-term re-do surgery (Table 6-12). Similar observations were found regarding very small (< 35 mm) or large (> 45 mm) femoral head

sizes. Within the prosthesis group, hip resurfacing was assigned a slightly higher weight compared to fixation type.

Revision

Rank	Feature importance (%)	Response graph
1	Age (31.6%)	
2	Prosthesis group (18.2%)	
3	BMI (11.2%)	
4	ASA (7.8%)	
5	Diagnosis group (7.6%)	
6	Femoral head size (6.0%)	

Re-operation

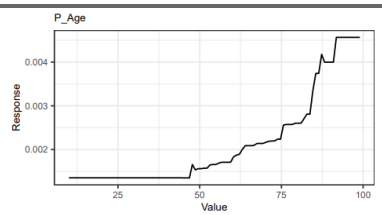
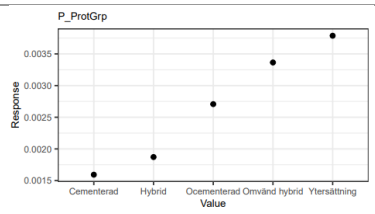
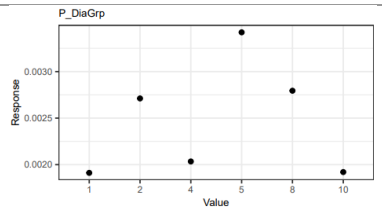
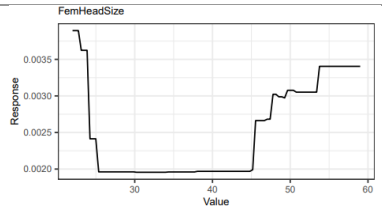
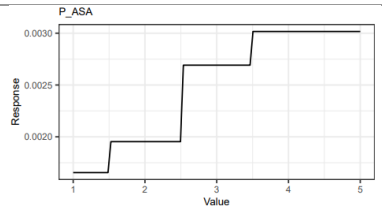
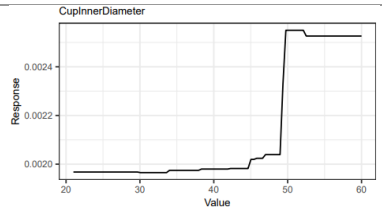
Rank	Feature importance (%)	Response graph
1	Age (21.0)	
2	Prosthesis group (18.3%)	
3	Diagnosis group (11.3%)	
4	Femoral head size (10.9%)	
5	ASA (9.0%)	
6	Cup inner diameter (5.3%)	

Table 6-12 Feature importance for 1-year revision (left) and re-operation (right) GBM models

6.4. Discussion

6.4.1. Summary of results

This study has demonstrated that machine learning-based predictive models for revision, and re-operation due to periprosthetic femoral fractures (PPFF) after primary hip arthroplasty achieved a relatively higher level of accuracy, as determined by the area under the curve (AUC), compared to the logistic regression linear model. To our knowledge, this is the first study to develop individualised risk assessment for PPFF.

Descriptive statistics of the SHAR database revealed no statistically significant difference temporal trends concerning the cumulative incidences of both revision and re-operation, suggesting consistency in these measures over the studied time period. While no gender discrepancies were apparent, the incidence of fractures seemed to be most pronounced in the 50 – 79 age groups. Although the recent SHAR annual report lacks specific data related to femoral periprosthetic fractures, it highlights that the number of revisions performed in the past decade was relatively constant with no significant differences based on gender (5). In contrast, the National Joint Registry for England and Wales' annual report of 2022, highlighted a decline in the incidence of revisions in the last decade compared to year 2008, and an inverse relationship between revision probability and the patient's age with greater variability among females at younger ages and among males at older ages (6). A New Zealand Joint Registry study showed a markedly greater lifetime risk of periprosthetic fracture following primary THR in males across all age bands, and particularly in the younger age groups (5.31%:2.82% in the 45 – 50 year age band) (415).

Our machine learning models showed good binary discriminative power in distinguishing between the positive and negative cases across various threshold settings, ranging between AUC 0.79 – 0.86 for both revision and re-operation predictions due to

femoral fractures. The Gradient Boosting Machine (GBM) was consistently the best performing model (except for 30-day revision; GBM AUC 0.84, LASSO AUC 0.85), and marginally outperformed the linear logistic regression model. Our model has a practical application in surgical practice and can be implemented either using a formula or a web calculator interface, in light of the limitations discussed in the following sub-section.

Two studies developed predictive models to forecast the risks of revision and re-revision following primary total hip arthroplasty (380, 381). Klemt *et al.* were able to achieve a discriminative ability of AUCs of 0.82 – 0.87 for all-case revision (380) and 0.80 – 0.86 for re-revision (381) following THR. However, their models were developed on data derived from a single tertiary referral centre in the United States; had a limited sample size ($n_{revision}$ = 566, $n_{re\text{-}revision}$ = 408); and did not include disease severity, pre-operative diagnosis, or implant-specific factors in their models – all of which were observed to be leading players in driving our predictions.

Our findings identified several patient characteristics and implant-specific features to be associated with revision and re-operation after THR due to PPFF at each time point. For the early (30 to 90 days) re-do surgery, these included:

- **Patient features:** increasing age and co-morbidities (ASA > 3); obesity classes II and III; and higher ASA grade (> 3) increased risks of both revision and re-operation.
- **Pre-operative diagnosis:** idiopathic necrosis increased revision risk.
- **Implant design:** cementless femoral fixation; hybrid fixation; hip resurfacing; and small (< 35 mm) or large (> 52 mm) femoral heads increased both risks.

In addition to the abovementioned, a pre-operative diagnosis of inflammatory arthritis was identified to have increased revision and re-operations risks at one year.

The risk factors identified in this study are in concordance with previous observational studies (214). Wright *et al.'s* systematic review of eighty-six studies found younger age, greater co-morbidity, pre-operative diagnosis of avascular necrosis, and the extremes of femoral head size all to be associated with increased revision risk (214). However, their findings with regards to prosthesis materials and fixation method, whilst favouring the use of cement, was inconclusive and this could be primarily attributed to the methodology employed which excluded studies of small sample sizes and those examining failure of a specific component, such as the stem. In contrast to the available evidence suggesting that youth predispose to exceeding the lifetime risk of revision surgery (227, 228), our study focused on a specific reason for short-term (up to one year) re-do surgery due to femoral fracture, thus, our augmented one-year implant failure risk with increasing age is likely attributed to the underlying aging physiological processes (see Section 1.9.1, (226, 416)). With regard to comorbid status and body mass index, our findings support the available evidence indicating increased odds of early revision THA with higher co-morbidity indices (277, 278) and BMI (see Section 1.9.1).

Our study showed that prosthesis choice, specifically cementless femoral fixation and hip resurfacing arthroplasty, seem to be the strongest risk factor for periprosthetic femoral fracture. That said, the limitations of using revision as an endpoint outcome when making comparisons between implants ought to be considered and are discussed in Section 1.8. The following chapter investigates the effect of implant fixation on mortality, revision and periprosthetic femoral fracture risks using a propensity score matching method.

6.4.2. Limitations and further work

The study findings should be interpreted in light of several limitations. First of all, the limitations related to the retrospective observational nature of the study using a single national joint registry as well as the limitations concerning the use of revision/re-

operation as endpoint outcomes, as discussed in the previous chapters, remain and hinder the clinical applicability to other settings. Similarly, our models were compared using the area under the curve (AUC) metric, which is regarded as a gold-standard in evaluating classifier performance across a range of thresholds and allow comparisons of multiple machine learning models. However, the AUC-ROC curve-based metric neglects to address the balance of true positives and true negatives, which alternative measures seek to address. However, the need to specify a threshold in other approaches is known to introduce variations in clinical practice. Additionally, the feature importance identified by our models whilst seeking to provide insights into the risk factors, are not subject to inferential testing assumptions, hence cannot establish associations. Future studies should aim to externally validate our machine learning models on other datasets, such as the NJR in the United Kingdom, and prospective clinical trials are required to investigate the effects of certain risk factors like co-morbidity indices (see Section 1.9) and fixation methods on patient outcomes.

6.4.3. Conclusions

This is the first study to develop machine learning models to forecast the risk of periprosthetic femoral fractures necessitating re-do surgery. Several risk factors were identified by our models, and these including: cementless fixation type; hip resurfacing arthroplasty; the extremes of femoral head sizes; a pre-operative diagnosis of idiopathic necrosis or inflammatory arthritis; and increasing age, BMI, and ASA co-morbidity grades. Our validated machine learning models on a large-scale national database showed good discrimination that estimates individuals' risk of femoral fracture and can be readily applied to surgical practice to aid clinical decision-making.

Chapter 7 The effect of implant fixation on safety, survival and periprosthetic fracture outcomes in primary elective hip arthroplasty

7.1. Introduction

Despite the prevalent use of cemented and cementless fixation in elective primary hip arthroplasty, the evidence behind the use of one fixation over the other is contentious (see Section 1.11). Cementless femoral fixation has been demonstrated in previous chapters to be a safer intervention with a lower mortality risk compared to cemented implants, however, its re-revision rate is higher. Chapter 4 employed Cox proportional hazards model with penalty regularisation coupled with multivariate analyses to investigate short- and long-term safety and implant survival following primary hip arthroplasty. Whilst the model is particularly useful for exploring complex high-dimensional predictor-outcome relationships, it might be influenced by unmeasured confounders. Despite implant survival being the predominant outcome measure in registry datasets, its utility is limited due to inherent constraints, as outlined in Section 1.8.2, leading to an ongoing debate among researchers regarding the risk-benefit assessment of cemented and cementless implants in primary hip arthroplasty.

As described in Chapter 6, the Swedish Hip Arthroplasty Register (SHAR) lacks specific information regarding the periprosthetic fracture risk comparison between cemented and uncemented fixation in primary elective hip arthroplasty. However, a previous study using the SHAR database between 1992 and 2007 reported an increased periprosthetic stem fracture risk in the use of cemented stems compared to cementless fixation (17%

versus 6%) (5, 345). This study is in line with the National Joint Registry of England and Wales (NJR), which shows an increased fracture risk following cementless and reverse hybrid fixation (0.67% and 0.66%) in comparison to cemented fixation (0.53%) (6), and the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) which highlights an age-group-dependent increased fracture risk for cementless implants (7). On the other hand, primary elective hip arthroplasty exhibits a consistently low and variable mortality rate across the three registries. The 90-day mortality rates in 2022 are reported to be 1.9% (SHAR) and 0.46% (NJR), respectively, (5, 6) while the AOANJRR reports a one-year mortality risk of 1.6% (7).

Most randomised controlled trials for primary hip arthroplasty targeted neck of femur fractures or congenital dysplasia. The single-centre RCT in literature for primary elective hip replacement among two surgeons followed 124 patients for 17 years and reported superior survivorship status for cementless implants, especially in younger patients (349). Observational and registry studies for elective primary hip arthroplasty vary in their susceptibility to bias and confounding by indication. These studies often have limited adjustments for patient characteristics, diagnosis, or surgical factors that may influence treatment decisions and outcomes. As a result of this, practice with respect to fixation in primary elective hip replacement varies widely among centres and surgeons. Debate continues as to whether cemented fixations yield superior safety and survival outcomes (347). Thus, it is important to carefully consider the balance between mortality and revision risks and the potential confounding factors.

This chapter aims to determine the effect of cementless femoral fixation on outcome, both in terms of safety and implant survival, using the Swedish Hip Arthroplasty Register database. Multiple outcomes will be investigated including mortality, revision, and

revision due to periprosthetic fracture between thirty days and two years. To mitigate the issue of confounding by indication, propensity score matching techniques have been employed to create matched groups of patients. The details of these techniques can be found in Section 7.3. Part two of this thesis uses a biomechanics approach to understanding strain behaviour, fixation, and fracture risk using cementless femoral implants.

7.2. Method

7.2.1. Data sources

The data sources used in this chapter are derived from the SHAR datasets, first described in Chapters 2 and 3, and encountered in Chapter 6. The variables used are displayed in Table 6-1.

7.2.2. Exposures and outcomes

Analyses were restricted to patients undergoing primary elective hip replacement. Patients who had undergone hip arthroplasty due to specific diagnoses such as trauma, tumour-related conditions, and degenerative conditions other than osteoarthritis were excluded. Patients below the age of 18 were excluded to minimise potential confounding factors that could arise from the different outcomes observed amongst paediatric patients (369-371).

Revision included patients who underwent an additional surgery to the hip where at least one component (cup and/or stem) is exchanged, extracted from, or added to the prosthesis. The failure rates of cemented and uncemented femoral fixation were compared

by three metrics: death, rates of revision and rates of revision due to femoral periprosthetic fractures, between thirty days and two years.

7.2.3. Statistical analysis

7.2.3.1. Propensity score matching

Propensity score matching techniques facilitate the creation of matched comparison cohorts to mitigate the influence of variation in baseline characteristics, minimise confounding by indication and enable a robust assessment of the treatment effect (417, 418). Unlike logistic regression for estimating the propensity scores, generalised boosted regression modelling (GBM) is a non-parametric machine learning technique capable of handling complex covariate interactions and non-linear relationships. GBM was employed, using the "twang" package, to calculate the weights for the comparison cases (cemented versus cementless), estimating the average treatment effect on the treated (ATT) over multiple iterations and interaction depths. The measured pre-treatment covariates in the weighted treatment (cementless fixation) and the control group (cemented fixation) were assessed for balance, or equivalence. The optimal number of GBM iterations, which minimises the differences in covariates between the treatment and control groups, was chosen based on the absolute standardised mean difference, often referred to as the effect size, using the mean of the balance metric across variables. The resulting estimates were used to compute a propensity score signifying the probability of each observation receiving a cementless femoral fixation. A one-to-one "nearest neighbour" matching approach with a narrow calibration width was implemented, where each cemented was matched to a single cementless femoral fixation based on the propensity score. Table 7-1 provides an overview of the confounders considered in the propensity score calculations.

Propensity score matching covariates	
Age	Body mass index (BMI)
Gender	Hospital type
ASA grade	Pre-operative diagnosis

Table 7-1 Covariates used for the propensity score estimation

7.2.3.2. Diagnostics and sensitivity analyses

Tabular and graphical balance diagnostics were used to evaluate inter-group baseline values, ensuring that the matching process was successful in creating comparable control and treatment groups. Tabular assessment included crude and matched summary statistics including mean, percentage, standardised mean difference (SMD), and p-value of chi-squared (for categorical covariates) or t-test (for continuous covariates). Graphical assessment included evaluating the distribution of control and treatment groups using histograms and boxplots. Given the balanced nature of baseline characteristics post-matching, unconditional univariable logistic regression models for the binary outcomes (mortality, revision, and periprosthetic femoral fracture) and Kaplan Meier survival models for 30, 60, 90 days, and 1 and 2 years were used. As before, analyses were performed using R statistical computing environment version 4.3.0. R packages used for the propensity score matching include: 'MatchIt' (version 4.5.3), 'survival' (version 3.5.5), 'twang' (version 2.5), and 'optmatch' (version 0.10.6).

7.3. Results

7.3.1. Crude summary statistics

A total of 154,519 SHAR records (101,523 cemented and 52,996 cementless) were available for the purpose of matching. Baseline crude variables are shown in Table 7-2.

Covariates	Crude		P value	SMD
	Cemented	Cementless		
Number of operations	101,523	52,996		
Age at surgery	72.1 (8.3)	59.9 (10.0)	<0.001	1.312
Sex (Male)	38,585 (38 %)	28,245 (53.2 %)	<0.001	0.311
Unit type			<0.001	0.395
University Hospital	5,895 (5.8%)	6,366 (12.0 %)		
County Hospital	31,798 (31.3 %)	15,220 (28.7 %)		
Rural Hospital	44,591 (43.9%)	15,624 (29.8 %)		
Private Hospital	19,239 (19.0%)	15,786 (46.7%)		
Body Mass Index (BMI)			<0.001	0.131
Underweight	956 (0.94 %)	242 (0.45 %)		
Normal	33,200 (32.7 %)	14,828 (27.9 %)		
Overweight	43,075 (42.4 %)	23,287 (43.9 %)		
Obesity Class I	18,713 (18.4 %)	11,025 (20.8 %)		
Obesity Class II	4,658 (4.5 %)	2,981 (5.62 %)		
Obesity Class III	921 (0.9 %)	633 (1.19 %)		
American Society of Anaesthesiologists grade			<0.001	0.436
1	1,7981 (17.7 %)	18,833 (35.5 %)		
2	63,727 (62.7 %)	28,304 (53.4 %)		
3 +	19,815 (19.5 %)	5,859 (11.0%)		
Pre-operative Diagnosis			<0.001	0.258
Primary arthrosis	94,752 (93.3 %)	46,572 (87.9 %)		
Inflammatory	1,301 (1.3 %)	661 (1.2 %)		
Childhood dx	910 (0.9 %)	2,351 (4.4 %)		
Idiopathic necrosis	2,510 (2.5 %)	1,230 (2.3 %)		
Secondary arthrosis	1,962 (1.9%)	2,140 (4.0 %)		
Other	88 (0.1 %)	42 (0.1 %)		

Table 7-2 Baseline crude variables pre-matching

There were statistically significant differences between the groups for all the covariates ($p < 0.001$).

7.3.2. Matching

983 GBM iterations were used to minimise the differences in covariates between the treatment (cementless) and control (cemented) groups, in other words, to optimise the mean effect size of the average treatment effect on the treated (Figure 7-1).

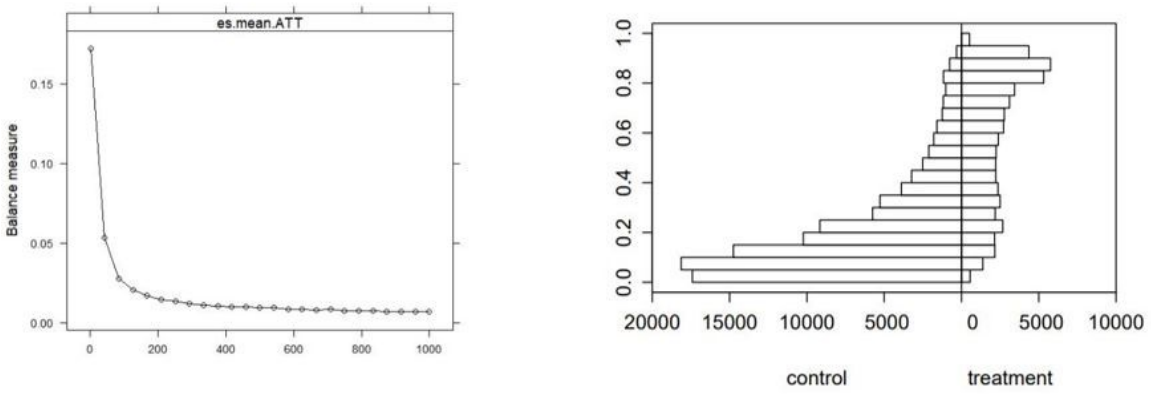


Figure 7-1 Propensity score matching process
Left – The effect size of average treatment effect on the treated (cementless fixation) using the maximum of the balance metric over 500 iterations. ***Right*** – Propensity scores for the control and treatment groups pre-matching

Distribution of propensity scores before and after matching are illustrated in Figures 7-1 to 7-3. Satisfactory graphical balance was achieved between groups. Statistical balance diagnostics were performed and are presented in the following sub-section.

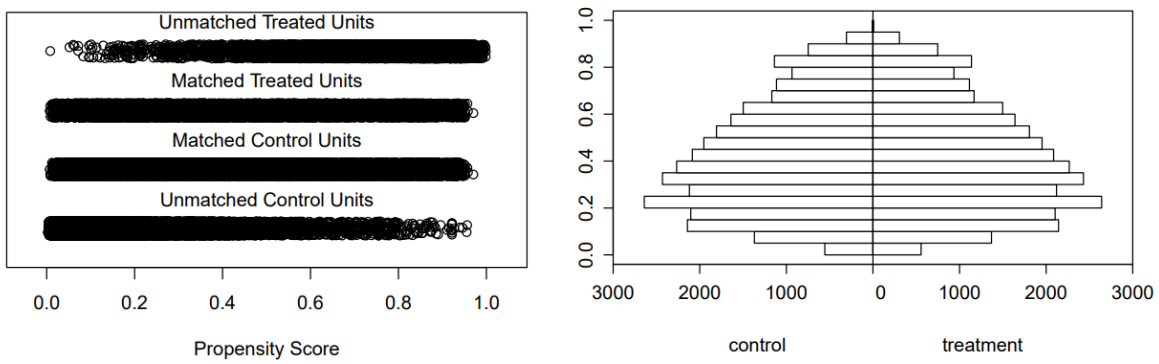


Figure 7-2 Propensity scores distribution post-matching
Left *– Distribution of propensity scores for unmatched and matched treatment and control groups,* ***Right*** *– Propensity score distribution for the control and treatment groups post-matching*

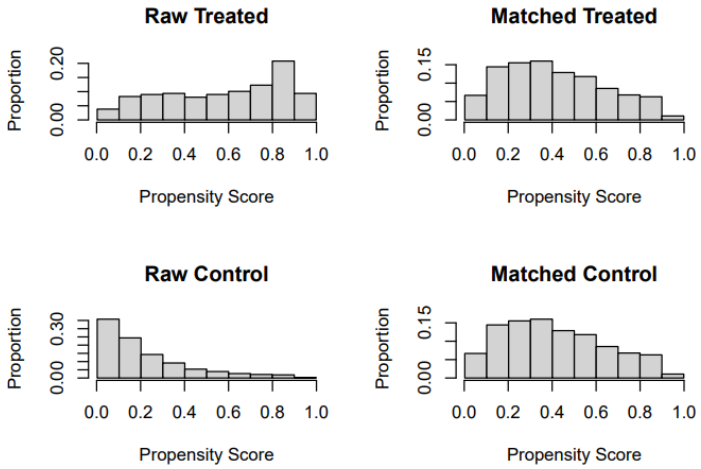


Figure 7-3 Distribution of crude and matched groups per propensity score

7.3.3. Matched summary statistics

Using the propensity scores, it was possible to one-to-one match 30,032 out of 52,996 (56.6%) cementless with cemented fixation. Baseline variables post-matching are presented in Table 7-3. Overall, there was no statistically significant differences between groups ($p > 0.05$). These patient cohorts were used to compare safety and implant failure outcomes in this study.

Covariates	Matched		P value	SMD
	Cemented	Cementless		
Number of operations	30,032	30,032		
Age at surgery	65.36 (7.88%)	65.29 (8.06%)	0.301	0.008
Sex (Male)	14,823 (49.4%)	14,833 (49.4%)	0.941	0.001
Unit type			0.134	0.019
University Hospital	2,727 (9.1%)	2,891 (9.6%)		
County Hospital	9,949 (33.1%)	9,921 (33.0%)		
Rural Hospital	10,160 (33.8%)	10,123 (33.7%)		
Private Hospital	7,196 (24.0%)	7,097 (23.6%)		
Body Mass Index (BMI)			0.619	0.015
Underweight	100 (0.3%)	122 (0.4%)		
Normal	8,681 (28.9%)	8,653 (28.8%)		
Overweight	13,309 (44.3%)	13,199 (43.9%)		
Obesity Class I	6,100 (20.3%)	6,192 (20.6%)		
Obesity Class II	1,554 (5.2%)	1,574 (5.2%)		
Obesity Class III	288 (1.0%)	292 (1.0%)		
American Society of Anaesthesiologists grade			0.415	0.013
1	7,727 (25.7 %)	7,754 (25.8%)		
2	18,079 (60.2%)	17,999 (59.9%)		
3 +	4,226 (14.1%)	4,279 (14.3%)		
Pre-operative Diagnosis			0.064	0.026
Primary arthrosis	28,236 (94%)	28,061 (93.4%)		
Inflammatory	270 (0.9%)	281 (0.9%)		
Childhood dx	472 (1.6%)	518 (1.7%)		
Idiopathic necrosis	441 (1.5 %)	483 (1.6%)		
Secondary arthrosis	606 (2.0%)	676 (2.3%)		
Other	7 (0.0%)	13 (0.0%)		

Table 7-3 Baseline variables post-matching

7.3.4. Comparison of safety and failures

Non-time-dependent unconditional logistic regression odds ratio showed that mortality was significantly less likely with cementless femoral fixation whilst revision and femoral periprosthetic fractures were significantly more likely to occur in the crude and matched

comparisons (Table 7-4). However, the effect on mortality was greater in the crude comparisons and vice-versa for the effect of periprosthetic femoral fracture.

	Crude comparisons		Propensity matched comparisons	
	Odds ratio	Significance	Odds ratio	Significance
Revision	1.51 (1.43-1.60)	< 0.001	1.51 (1.38-1.65)	< 0.001
Periprosthetic femoral fracture	1.57 (1.37-1.81)	< 0.001	2.40 (1.92-2.99)	< 0.001
Death	0.28 (0.26-0.29)	< 0.001	0.73 (0.69-0.78)	< 0.001

Table 7-4 Crude and matched logistic models
Odds ratios <1 favour cementless femoral fixation

7.3.5. Survival comparison

At two years, the unadjusted hazard ratio for mortality for cementless femoral fixation was HR 0.82 (95% CI 0.71-0.94, p = 0.006) compared to the cemented group (Figure 7-4). Whilst statistically significant at two years, the unadjusted hazard ratio for short-term mortality outcome was not significant between the two groups (Table 7-5).

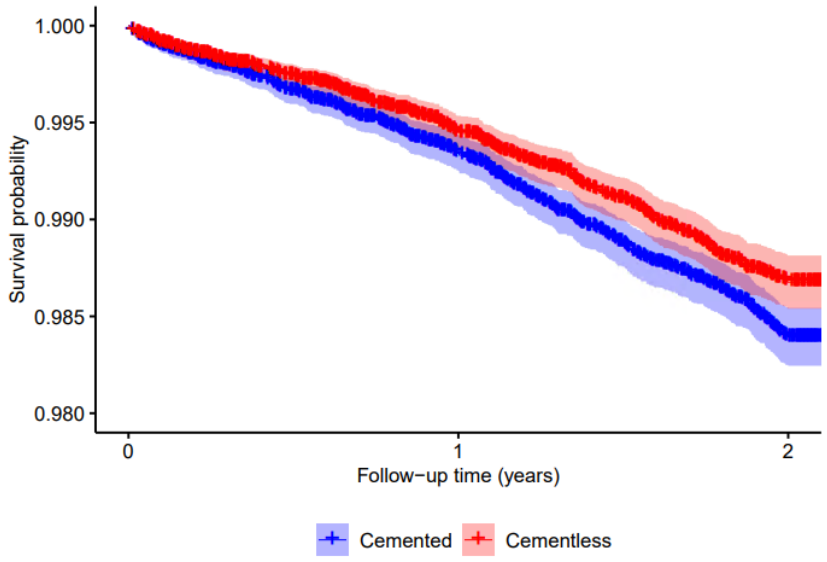


Figure 7-4 Mortality rates of cemented and cementless femoral fixation

Outcome	Time-point	Survival (Cemented)	Survival (Cementless)	Cox HR	Significance
Revision	30 days	99.6 (99.5-99.7)	99.1 (99.0-99.2)	2.16 (1.73-2.68)	< 0.001
	90 days	99.4 (99.3-99.5)	98.6 (98.5-95.8)	2.27 (1.90-2.71)	< 0.001
	1 year	99.0 (98.9-9.91)	97.9 (97.8-98.1)	2.14 (1.85-2.46)	< 0.001
	2 years	98.7 (98.5-98.8)	97.2 (97.0-97.4)	2.07 (1.83-2.34)	< 0.001
Revision due to femoral periprosthetic fracture	30 days	100 (100-100)	99.7 (99.7-99.8)	41.1 (10.1-167.2)	< 0.001
	90 days	100 (100-100)	99.6 (99.5-99.7)	17.5 (8.18-37.5)	< 0.001
	1 year	99.9 (99.9-100)	99.5 (99.4-99.6)	6.88 (4.35-10.89)	< 0.001
	2 years	99.9 (99.9-99.9)	99.5 (99.4-99.6)	5.54 (3.70-8.29)	< 0.001
Death	30 days	99.9 (99.9-99.9)	99.9 (99.9-100)	0.88 (0.50-1.55)	0.70
	90 days	99.8 (99.8-99.9)	99.8 (99.8-99.9)	0.80 (0.54-1.19)	0.30
	1 year	99.3 (99.2-99.4)	99.4 (99.4-99.5)	0.83 (0.67-1.03)	0.09
	2 years	98.4 (98.2-98.5)	98.7 (98.5-98.8)	0.82 (0.71-0.94)	0.006

Table 7-5 Matched survival models at two years
Hazard ratios <1 favour cementless femoral fixation

In matched patients, revision was more prevalent in the cementless group (Table 7-5). Unadjusted hazard ratios for revision for the different time points were statistically significant ($p < 0.001$) and ranged between HR 2.07 (1.83-2.34) at two years, and HR 2.27 (1.90-2.71) at 90 days (Figure 7-5). At two-years, overall implant survival was 98.7% (95% CI 98.5-98.8) for cemented and 97.2% (95% CI 97.0-97.4) for cementless fixation. Incorporating all case revisions diminished the overall survival statistics and attenuated the difference between cemented and cementless cohorts.

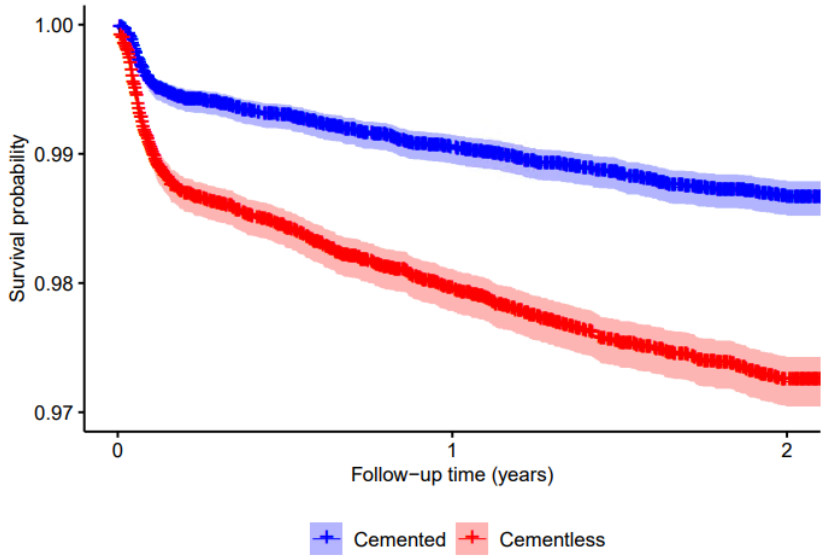


Figure 7-5 All-case revision rates of cemented and cementless femoral fixation

When revisions due to femoral periprosthetic fracture were studied, the effect was augmented and still significant, particularly in the short interval postoperatively (30-day HR 41.1, 95% CI 10.1-167.2, $p < 0.001$, Figure 7-6). There were more femoral periprosthetic fractures for the cementless group in the first 90 days post-operatively (HR 17.5, 95% CI 8.18-37.5), before the hazard plateaued (Table 7-5 and Figure 7-6).

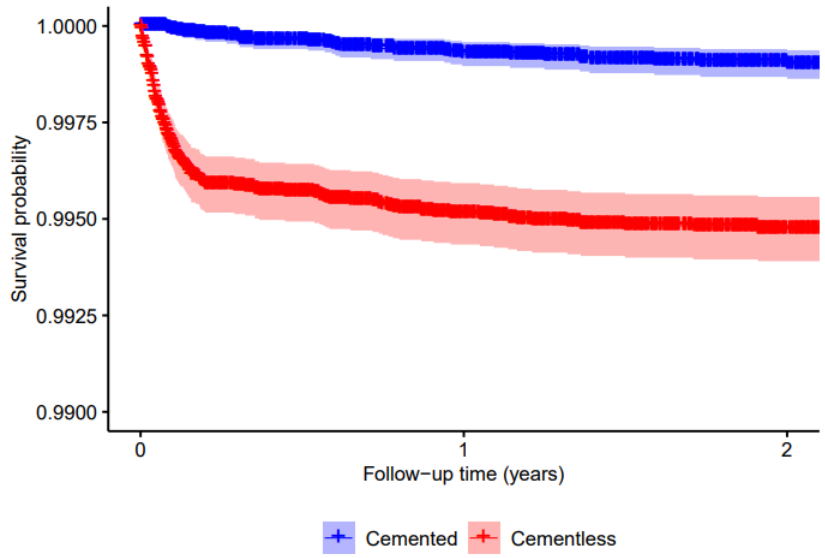


Figure 7-6 Rates of revision due to periprosthetic femoral fracture of cemented and cementless femoral fixation

In-depth sensitivity analyses with broader calibre widths consistently led to the same conclusions across all outcome measures.

7.4. Discussion

7.4.1. Summary of results

This study has demonstrated the effect of femoral fixation type in primary elective hip arthroplasty on mortality and implant failure outcomes. Here we showed, similar to our findings in Chapter 4, that in a matched cohort of individuals of similar baseline

characteristics, cementless fixation has a significantly lower two-year mortality risk but higher incidences of short-term femoral periprosthetic fracture and all-case revision.

Patients who undergo cemented femoral fixation appear to have an overall 27% increased risk of mortality in which 18% of that is likely due to an increase at the two-year interval postoperatively. Previous studies highlighted superior safety outcomes of uncemented femoral implants (see Chapter 1.11; (346, 419-421)). A retrospective Nordic Arthroplasty Register study reported higher mortality rates in the crude comparisons between cemented and cementless fixations at 30-day (Crude HR 1.7, 95% CI 1.3–2.1) and 90-day (HR 1.6, 95% CI 1.4–1.9) postoperatively, favouring cementless fixation (422). However, when they adjusted for age, gender, year of surgery, and comorbidity, they found no statistically significant differences (30-day HR 0.94, 95% CI 0.71–1.3; 90-day HR 0.97, 95% CI 0.79–1.2). Notwithstanding the absence of adjustments for variables such as body mass index (BMI), hospital type, and pre-operative diagnosis, their outcomes align with our own, which similarly did not demonstrate statistical significance at earlier time points. In contrast, Garland *et al.* matched individuals who underwent either cemented or cementless fixation for total hip arthroplasty (146,818 cemented) with 862,294 unexposed individuals from the general population and reported a statistically significant increased risk of early death in patients receiving cemented THA that is reversed from two weeks postoperatively, but not in the cementless-control groups, leading the authors to suggest that cemented femoral fixation is likely associated with higher incidences of thromboembolic events (348). While it can be contended that the association between to the fixation method and the occurrence of mortality diminishes as time elapses since the primary THR (423), the augmented two-year mortality risk in our matched cohorts

warrants further investigations into the factors leading to a time-dependent impact (see Chapter 4).

In terms of revision and femoral periprosthetic fracture, uncemented fixations were shown to be significantly associated with increased risks at all time points (all-case revision OR 1.51, 95% CI 1.38-1.65; revision due to fracture OR 2.40, 95% CI 1.92-2.99), and particularly in the short-term interval postoperatively, which harmonises with findings from earlier studies in the literature (see Section 1.11 and (422)). This suggests that when selecting fixation method, one ought to account for the heightened revision and fracture risks associated with cementless fixations as well as the mortality associated with such secondary procedures.

7.4.2. Limitations and further work

Whilst the observed difference between cemented and cementless femoral fixations are partially explained by diagnostic, patient, and surgical factors that were matched in this study, our findings alluded to a residual difference in survival between the fixation groups. Other factors are likely to contribute to the calculated hazards, including patient factors (specific co-morbidities, stage of disease, and socioeconomic and lifestyle variables), pre- and post-operative care (the use of mechanical or chemical thromboprophylaxis), and surgical factors (surgical expertise and technique). However, we lacked access to such information in its entirety. If the use of thromboprophylaxis, or surgical expertise, are shown to be related to the fixation method, our findings would be confronted with residual confounding. Future studies also need to explore the effect of fixation type on patient-reported outcome measures and assess pre- and post-operative radiographs.

Likewise, our findings are constrained to the studied Swedish population wherein most fixations are cemented. These may bear varying results if assessed using different populations such as the National Joint Registry of England and Wales. Furthermore, we implemented a matching protocol with a narrow calibre width to attain a high level of covariate balance, acknowledging the trade-off of reduced sample size and treatment effect precision. Nonetheless, our sample size remained sufficiently robust to detect meaningful differences, and our sensitivity analyses, which incorporated wider calibre widths, consistently supported our conclusions.

7.4.3. Conclusions

The evidence from this study indicates the potential for improved long-term safety outcomes associated with uncemented femoral implants, albeit with a trade-off of increased risks of revision and periprosthetic femoral fracture. The underlying reasons for this discrepancy are multifaceted and not entirely accounted for in the SHAR dataset. As explored in Chapter 4, patient factors (age, BMI, co-morbidities, and gender); pre-operative diagnosis; and unit type appear to play a significant role in this phenomenon. However, further investigations are warranted to understand the effects of competing variables such as: surgical technique and expertise; pre- and post-operative care; and lifestyle and socioeconomic factors.

Summary, Part One: Machine Learning & Outcomes following Primary Hip Arthroplasty

The findings from Part I of this PhD work shed a light on the importance of implant fixation method (cemented versus cementless) as a significant modifiable prognostic factor. Chapter 4 highlighted that cemented fixation appears to be associated with higher mortality rates, but lower rates of implant failure compared to cementless fixation. This is in agreement with prior literature on the Swedish and Norwegian registries (348, 378). The limitations of the study in Chapter 4, of not evaluating the reasons for revision nor adjusting for various co-variates could have augmented the effect of confounding by indication, informed Chapter 6 to predict the risk of periprosthetic femoral fracture necessitating re-do surgery, and Chapter 7 to investigate the effect of implant fixation on safety and implant failure outcomes using a propensity score matching technique.

Again, the findings from both chapters highlighted that uncemented femoral implants are associated with an increased risk of periprosthetic femoral fracture and revision risks. The machine learning models in Chapter 6 identified cementless implants as a driving feature for increased revision and re-operation risks necessitating an intervention, whilst Chapter 7 reported in a matched cohort of cemented and cementless fixations that uncemented femoral implants is associated with an improved long-term safety profile but a statistically significant increased risk of revision and periprosthetic femoral fracture. These findings are in agreement with previous literature (346, 348, 419-422) and with reports from the joint registries in Australia, Sweden and the United Kingdom showing that periprosthetic femoral fractures are the third most common cause for revision in THR. Likewise,

previous literature alluded to the fact that there remains heterogeneity in surgical technique and experience for cementless stem impaction and implant seating, wherein the impaction energy must be sufficient enough to seat the broach/implant fully but not excessively as to result in fractures (424, 425).

Part II of this thesis provides directional research translation, from the bedside (Part I) to the bench (Part II). It aims to address this question and enhance our understanding of this modifiable surgical technique of cementless femoral implant seating and fixation through two major biomechanical laboratory studies. Chapter 8 will use a digital image correlation technique and a validated rig that accounts for soft tissue compliance to measure strain behaviour in the proximal femur using synthetic bones of low and normal mineral densities. Chapter 9 will further test these findings in cadaveric "healthy bone" specimens using rosette strain gauges, providing guidance for further bench-side experiments in the field.

Results, Part Two: Optimising cementless short-stem femoral fixation using biomechanical approaches

Chapter 8 The effect of impaction energy and bone quality on cementless femoral implant fixation and femoral periprosthetic fracture risk

8.1. Introduction

Total hip arthroplasties are effective orthopaedic surgeries, typically indicated in patients with end-stage osteoarthritis (426). Cementless fixation for the femoral stem became popular in the late 1980s and has become the fixation of choice in Australia, Canada England and Wales, and Italy, (427). However, there remains controversy about whether cemented or cementless fixation provides superior outcomes. Cementless fixation can be more expensive than cemented, but this is offset by a shorter operative time. Cementless fixation also has a higher risk of periprosthetic fracture owing to the press-fit required to achieve primary stability of the stem. If the risk of periprosthetic fracture can be reduced, it will improve the outcome of cementless stems.

Initial cementless fixation relies on sequential impaction strikes to provide press-fit stability. Long term stability is provided by bone ingrowth, achieved via good initial bone and implant apposition (347). Appropriate impaction technique involves two key stages; a broaching step to prepare the intermedullary canal and a femoral implant insertion step, both of which are crucial to ensure satisfactory long-term osseointegration occurs between the cementless implant and the bone (428). The impaction energy must be sufficient enough to seat the broach/implant fully but not too excessive as to result in fractures (424). A previous study investigated the effect of strike technique on the seating and stability of

the cementless acetabular component, assembling femoral head and influencing tapper-trunnion stability (429-431).

Currently, surgeons rely on proprioceptive estimations of ideal impaction energy and number of strikes to achieve an optimal balance between stem/broach seating and fracture risk (432). An understanding of the biomechanical properties of a human femur may provide empirical evidence to improve fixation and minimise fracture risk. Strain measurement techniques can reliably predict regions of high stress in the bone during impaction. Cummins *et al.* (433) calculated the threshold force required for femoral impaction grafting using strain gauges, which was further validated by Flannery *et al.'s* findings (434). Strain gauges, however, only allows for discrete point measurements to be obtained (435), making it challenging to measure strain over the full duration of impaction. Non-contact optical digital image correlation (DIC) can give detailed directional strain magnitudes on three dimensional surfaces and is suitable for complex irregular structures such as femoral bone models (436, 437). DIC captures sequential images of a specimen during mechanical testing and computes deformation, strain, and displacement through analysis of spatial correspondences (438, 439). Previous studies used DIC to measure failure in proximal femurs under quasi-static loading conditions (440, 441), but if applied to the dynamic impaction process could yield new understandings of how strain fields are generated.

The aim of this study is twofold, firstly we propose a new method combining DIC with high-speed cameras to capture the rapid changes in bone strain during femoral impaction. Secondly, we investigate strain behaviour, fracture risk, and implant fixation in synthetic bone models during cementless femoral broaching and stem impaction. In other words, do higher impaction strikes lead to higher fracture risks or better fixation outcomes?

8.2. Materials and Methods

8.2.1. Specimen preparation

In this study, a total of 20 synthetic bones were used: 12 healthy synthetic specimens and 8 osteoporotic. All Sawbones® models were adult left femurs and manufactured by Sawbones® (Pacific Laboratories, Malmö, Sweden). The healthy and osteoporotic specimens were composed from 30 PCF and 15 PCF respectively thereby representing normal and low bone density. All specimens, healthy and osteoporotic, were prepared using a standard THA direct anterior approach (442). Instruments for cementless THA (Furlong Evolution, JRI Orthopaedics, Sheffield) were used to prepare femurs for impaction and during impaction testing itself.

Using an oscillating saw, each specimen was transversely cut at 17cm proximal to the intercondylar fossa. The mid-point between the lesser trochanter and the femoral head was identified using a resection guide, the neck was then resected at 45° to the long axis of the femur. To gain access to the medullary canal, a box osteotome was used to remove a section of cancellous bone from the upper-lateral aspect of the femoral neck (442).

To open the medullary canal a T-handled taper reamer was used to prepare the canal for broaching, with care taken to ensure that the reamer was aligned with the femoral shaft axis. The reamer was used until the minimum depth, as indicated on the instrument, had been met.

To prepare the intramedullary canal for receiving the femoral stem implant, the smallest sized broach (size 6) was attached to the broach handle and repeatedly hammered in and out to remove the synthetic cancellous bone. The broach was impacted until the proximal

part of the broach handle was aligned with the osteotomy level. To avoid varus positioning, a posterior lateral pressure was applied to the broach handle during impaction.

8.2.2. Application of the speckle pattern

For the DIC system to accurately capture data, a specially prepared high-contrast speckle pattern is applied to the surface of interest (443). DIC records and tracks the movement of the speckle pattern during mechanical testing, thus allowing surface strain to be computed (444). To achieve the speckle pattern, white and black paints were applied to each specimen two hours prior to impaction. In a well-ventilated space, the anterior surface of the femur was sprayed with a coat of white paint (VHT SP101) and set aside to dry for twenty minutes (445). Black spray paint (VHT SP102, Sherwin-Williams, California) was then used to apply the speckles by lightly pressing the trigger to achieve the required pattern. All specimens were sprayed from a distance to ensure only thin layers of paint were applied. GOM Correlate Professional, a post-processing software, was used to assess pattern quality prior to testing (440).

8.2.3. Set-up

When a specimen was ready for testing, it was placed into a custom 3D printed holder and secured onto the impaction platen (Figure 8-1) in a vertical loading axis at a 7° tilt (446).

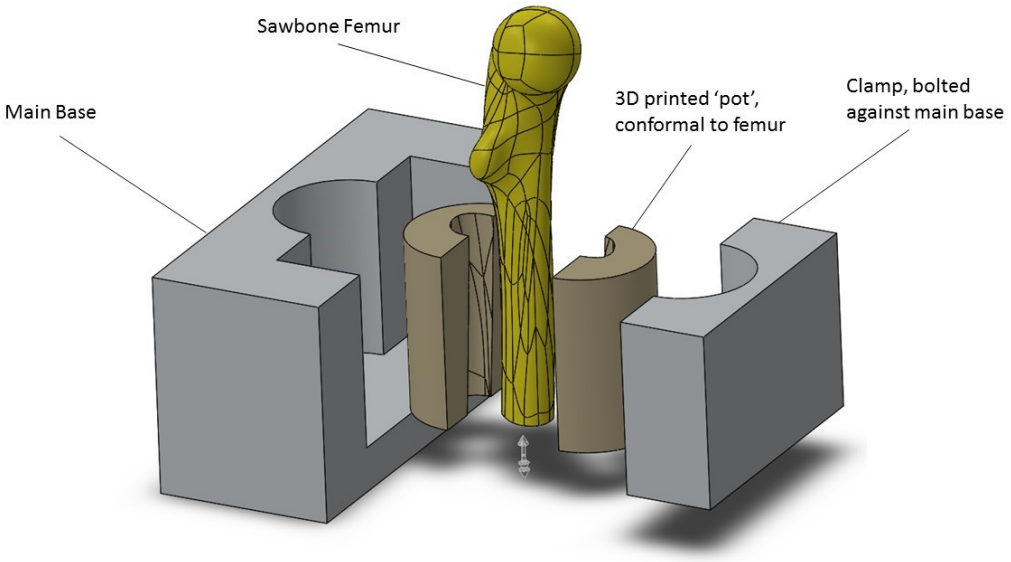


Figure 8-1 Assembly of platen and bone onto the rig

As seen in Figure 8-2, the impaction platen is secured to a spring/dashpot support. This provides the same displacement and acceleration rates as would be observed intraoperatively, thus accounting for soft tissue compliance and movement of the human femur (447). Strikes were delivered to each specimen by releasing the drop platen onto a secured broach handle below and subsequently the specimen. The rig drop platen was used to deliver impaction strikes to the specimen. A platen drop weight of 1.2 kg was chosen, representing the average mass of a THA surgical mallet. To vary the energy at which the specimens were impacted, the platen was released from pre-determined heights corresponding to specific energies.

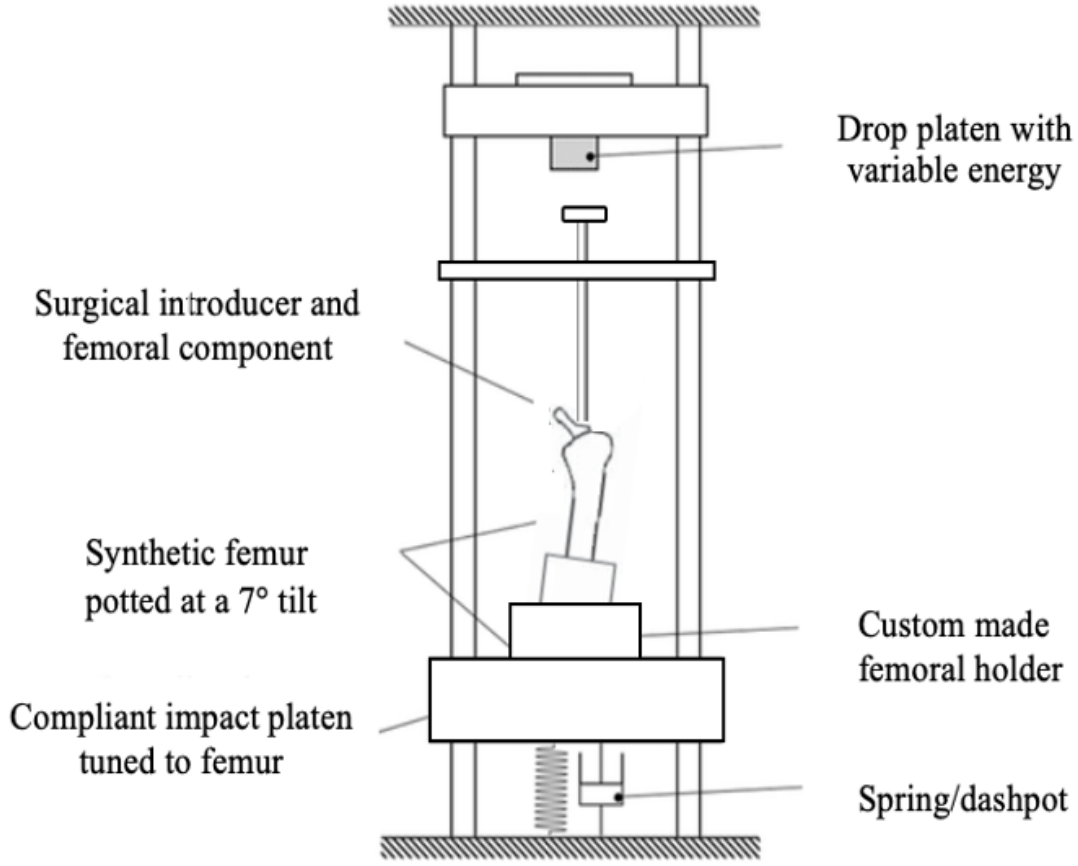


Figure 8-2 Experimental set up illustrating the in-vitro drop rig used for impaction (444, 448)

8.2.4. Testing

Using the in-vitro drop rig (Figure 8-2), femoral broaches of increasing size (7, 8, 9 and 10) were impacted into the femoral canal of each specimen. Beginning with a size 7 rasp, consecutive strikes were applied to the broach handle by releasing the drop platen to repeatedly impact the rasp into the femur. Consecutive strikes were applied until the rasp was seated. Broaches were hammered out of the femoral canal following every third strike. Following withdrawal, the broach was cleared of fragments and a further 3 strikes were then applied, with this process continued until the broach was seated.

The broach was considered seated once the cutting teeth were parallel to the depth of the neck resection, at which point the broach was hammered out, removed from the handle, replaced with the next size and the impaction process re-started. The broaching process was repeated, sequentially increasing the broach size until size 10, which was the size of the selected implant.

The intramedullary canal was deemed prepared for receiving the cementless stem prosthesis when the final broach, size 10, had been adequately seated and disengaged from the canal. A size 10 JRI Furlong femoral stem implant was then attached to an introducer and engaged into the femoral canal by hand. The introduction of the stem was completed by impacting the introducer until a tight metaphyseal fit was achieved (442).

Impaction energy levels were divided into three categories: low, medium, and high. Each energy level was defined based on measurements recorded in a study by Doyle *et al.* (448). In that study, a consultant orthopaedic surgeon simulated three levels of impact to the cadaveric specimens for both acetabular and femoral components; one mimicking the force used for patients with healthy bones, another mimicking the gentleness employed for patients with osteoporotic bones, and an intermediate impact (448). Three energy levels (J) were recorded E_{Low} = 1.26; $E_{,Medium}$ = 4.84; and E_{High} = 10.78.

Given that energy (J) is the product of mass (platen drop, kg) and the square of the velocity, energy levels generated for our study used a standard "mallet" weight of 1.20 kg and varied velocities (Table 8-1). The maximum velocity achieved by the rig was 4.08 (m/s), therefore, the high energy level in our study was 9.99 (J). On the healthy synthetic bone specimens, this protocol was performed at low, medium and high impaction energies. Whilst for the osteoporotic specimens the protocol was only performed at low

and medium energy levels since high velocity fractured low BMD sawbones during our pilot testing.

Energy Level	Low	Medium	High*
Drop Mass (kg)	1.20	1.20	1.20
Drop Velocity (m/s)	1.45	2.84	4.08
Energy (J)	1.26	4.84	9.99

Table 8-1 The impaction energy levels investigated: low, medium and high
Only healthy bones were impacted at this energy level

Twenty synthetic bones were tested totalling 600 impaction strikes; 12 normal BMD specimens were impacted at low, medium and high energy levels; and 8 low BMD samples were impacted at low, and medium energy levels (n=4 per energy level). As a result of the large number of strikes required to seat both broaches and implants at low and medium velocities, data analysis occurred for every third strike (i.e. strike 3, strike 6, strike 9…), and the average peak strain was calculated across velocities.

8.2.5. Digital image correlation (DIC)

To measure strain during mechanical testing images were captured using a three-dimension DIC system (441, 449). Two high-speed cameras (Phantom Miro) were set up as shown in Figure 8-3 (450). Both cameras were secured to a custom-made stand and rotated internally at a 12.5° angle. The impaction process was captured at 3,268 frames per second at a resolution of 768 x 768 pixels. The DIC system was calibrated prior to testing using a calibration model to establish the event volume. To increase specimen visibility during testing, additional lighting was used to illuminate the synthetic surface pattern (451). For osteoporotic bones each strike was analysed for peak strain measurements in regions representing Gruen zones 1, 2, 6 and 7. On the 12 healthy bones, the peak strain measurements were measured at regions representing Gruen zones 1, 2, 3, 5, 6 and 7.

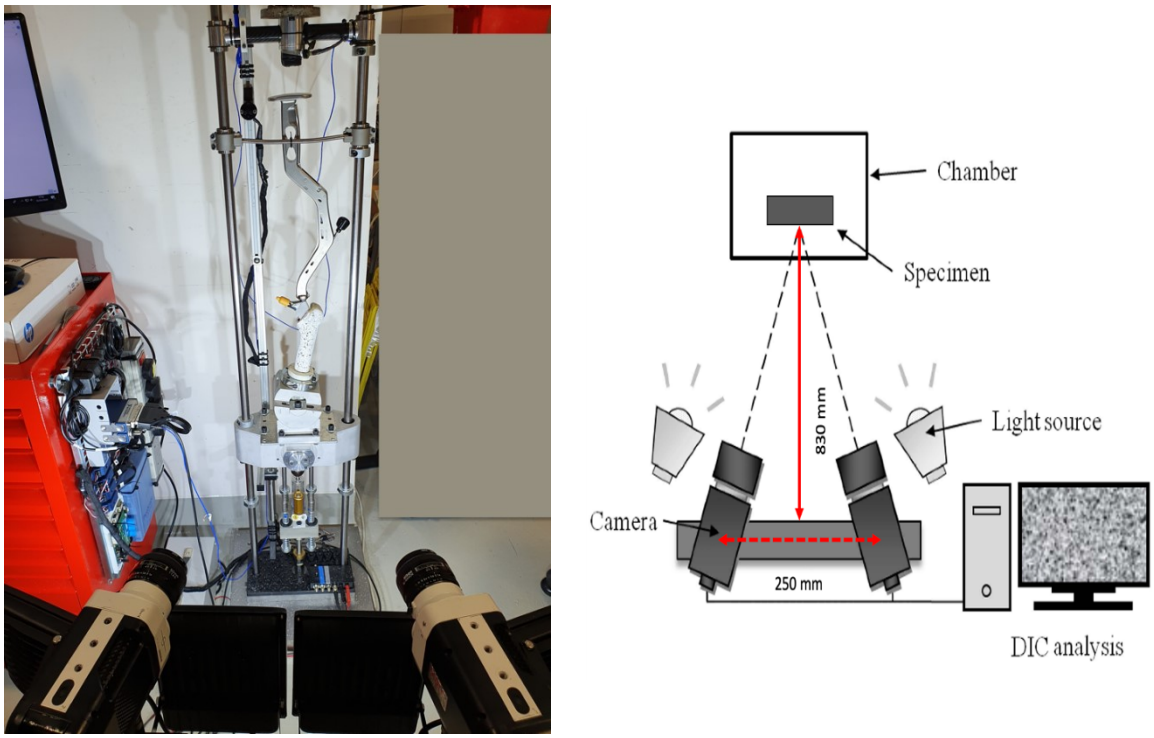


Figure 8-3 3D-DIC System setup
On the left a photograph taken by the author, and on the right a schematic of the 3D-DIC setup done by Kim *et al.* (452)

8.2.6. Fixation strength

To evaluate the initial stability of the femoral stem implant, pull-out fixation tests were conducted on a uniaxial loading testing machine (Instron model 5565; Instron, High Wycombe, UK) (Figure 8-4) (453). The femoral holder was secured to the test frame so that the actuator motion of the Instron was perpendicular to the axis of the femur. Implants were pulled out using an 11mm diameter steel rod which fixed distally to the implant and proximally to the machine actuator head. Tensile load was applied at a crosshead displacement rate of 0.2mm/min until the implant dislodged from the specimen (454).

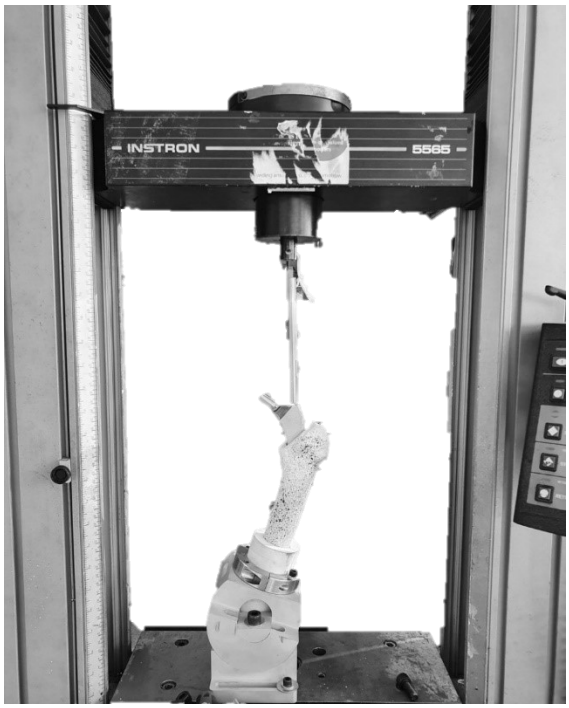


Figure 8-4 Pull-out test set-up using a uniaxial Instron

8.2.7. Statistical analysis

Variations in strain data between low, medium, and high velocity strikes in normal and low BMD specimens were evaluated for statistical differences. A two-way ANOVA determined statistical significance for normally distributed data whereas the equivalent non-parametric Friedman test was used for non-normally distributed data. The level of significance was established as $\alpha = 0.05$ for all tests. All analyses were performed using SPSS Version 26.0 and Microsoft Excel Version 22.0.

8.3. Results

For both normal BMD and low BMD specimens, a greater number of strikes was required for the low velocity versus either medium/high velocities for both broaching (two-way ANOVA; $p=0.0015$, $p=0.043$) and seating an implant (two-way ANOVA; $p<0.0001$, $p=0.013$, respectively). For normal BMD, no difference was detected regarding the number

of strikes between medium and high velocities for broaching (p = 0.88) and implant seating (p = 0.56). Figure 8-5 shows the mean number of strikes during mechanical testing for both specimen types at each impaction velocity.

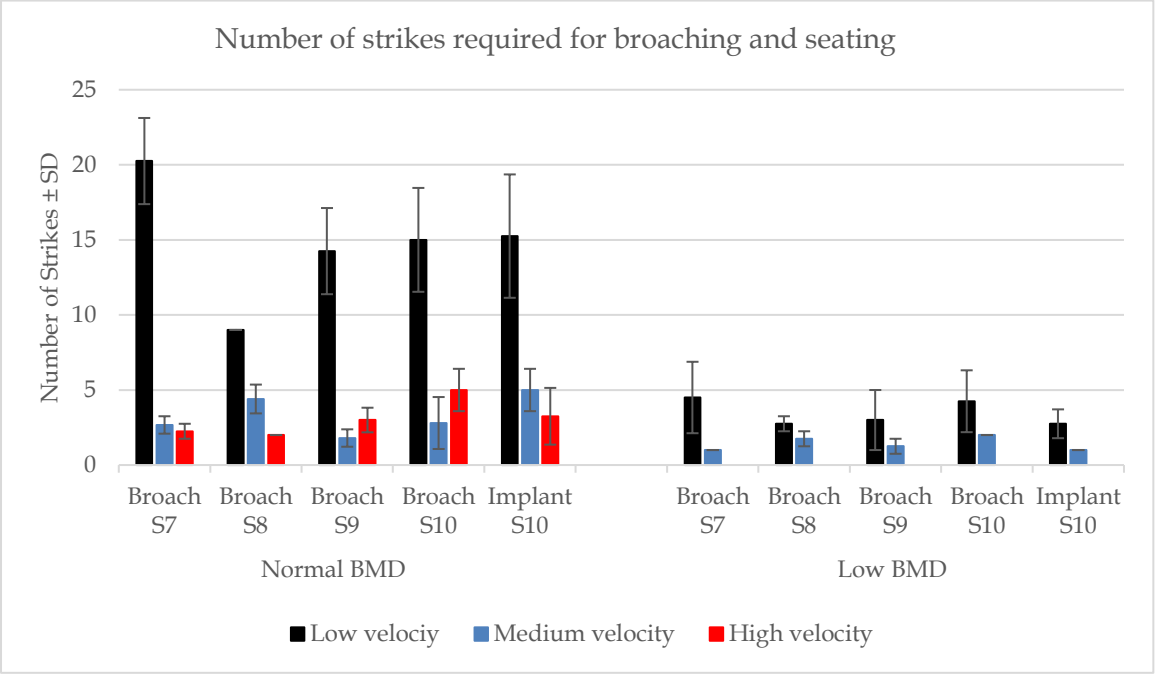


Figure 8-5 Average number of strikes required to seat broaches 7-10 and a size 10 implant at different mallet velocities in both normal BMD (healthy) and low BMD (osteoporotic) bones

Figure 8-6a represents a contour plot of DIC strain for a normal BMD specimen pre-impaction. Figure 8-6b represents a contour plot of DIC strain for a normal BMD specimen undergoing implant insertion with 9.99 energy (J) / 4.08 velocity (m/s). The pattern of strain is shown on the anterior proximal femur and the contour plot goes from -100.000 [%/s] to 100.000 [%/s]. Figure 8-6c is representative of a general pattern of strain observed during impaction in bone samples across all impaction velocities in Gruen Zones 1, 2, 6, and 7 in the direction of highest deformation. As shown in Figure 8-6c, after the strike an initial peak compressive strain was followed by a tensile strain. This was further followed by a post-strike tensile strain which diminished in magnitude prior to stabilising.

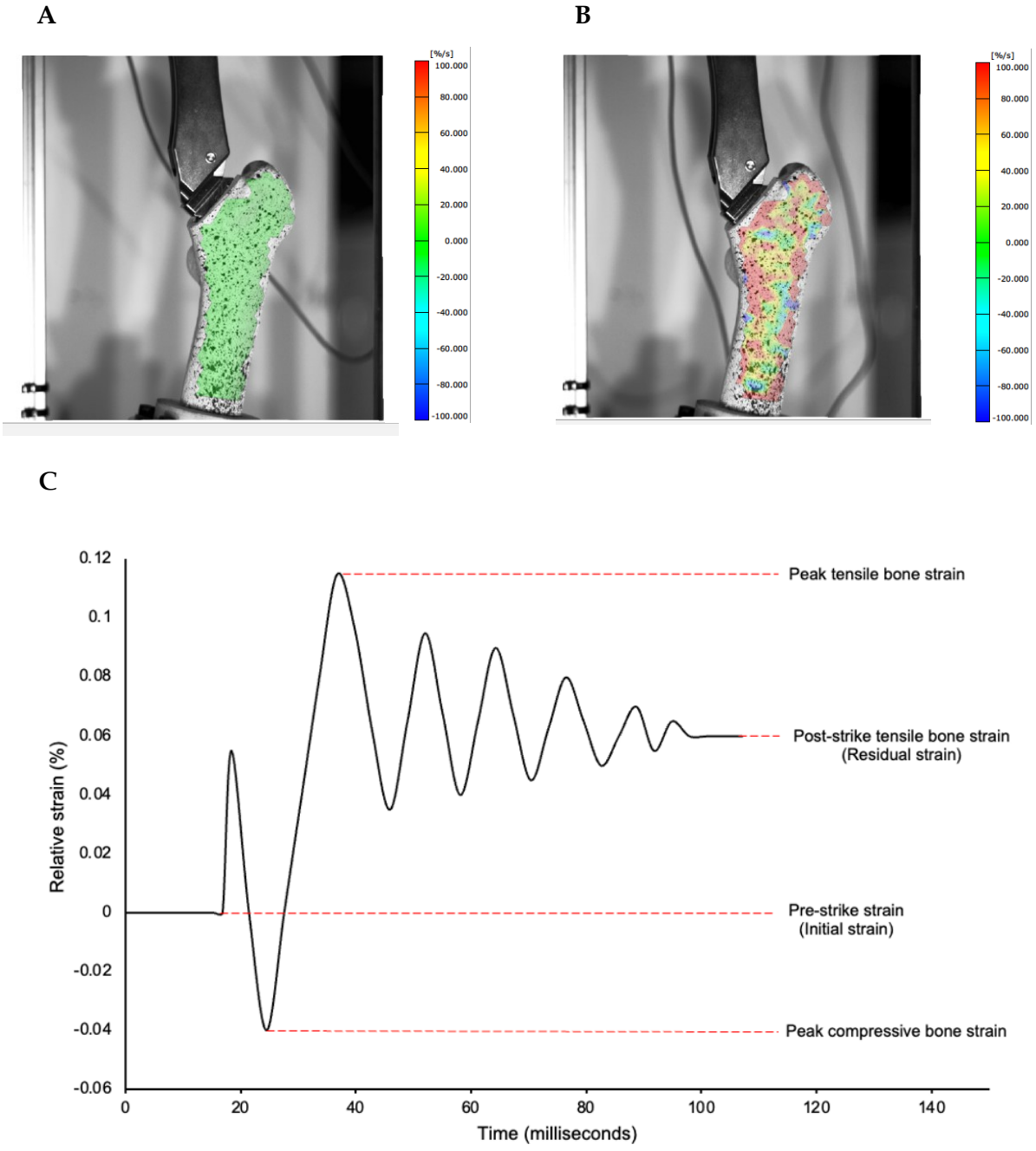


Figure 8-6 Illustrative contour plot pre-impaction (8-6a), during broaching impaction (8-6b) and (8-6c) showing strain behaviour during impaction from a bone sample

For normal BMD specimens, the average peak strain increased with each impaction velocity in the lateral Gruen zones (low vs medium $p<0.001$; medium vs high $p<0.05$; and low vs high $p<0.001$) (Figure 8-7a). For the medial Gruen zones the average peak strain increased for medium velocity compared to low velocity (low vs medium $p<0.05$) but did

not increase further for high impaction velocity (Figure 8-7a). In low BMD samples, there was no difference in peak strain for the different velocities in either the lateral (p=0.080) or medial (p=0.833) side (Figure 8-7b).

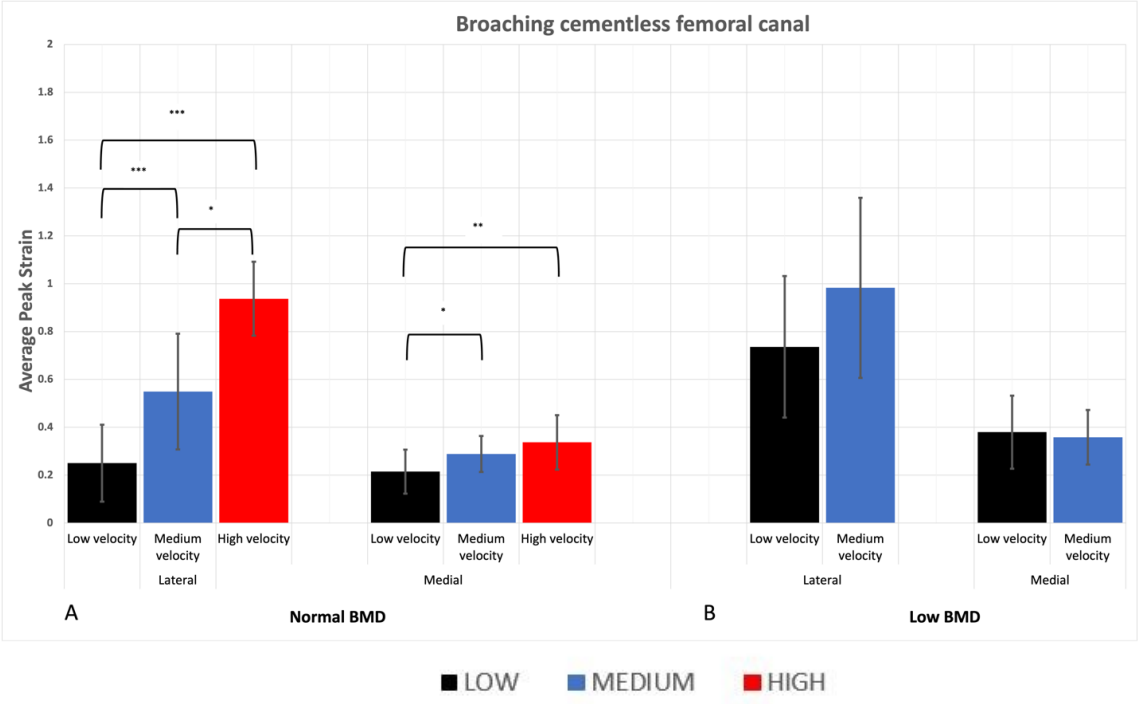


Figure 8-7 Average peak strain to broach a cementless femoral canal in normal (8-7a) and low BMD bones (8-7b) *p<0.05; **p<0.01; *<p<0.001**

For normal BMD specimens, there was no difference in the force required to remove specimens impacted by low or medium velocity strikes (p=0.604) with a mean of 685 N and 553 N, respectively (Figure 8-8). Similarly, there was no statistically significant difference between the force required to remove specimens impacted by high velocity strikes (mean = 1091 N) in comparison to low velocity strikes (p = 0.169). For the low BMD specimens, there was no difference in the force required to remove the stem from specimens impacted with either impaction velocity (p = 0.229). There was a clear difference between the force required to remove the stem from low BMD specimens and normal BMD specimens.

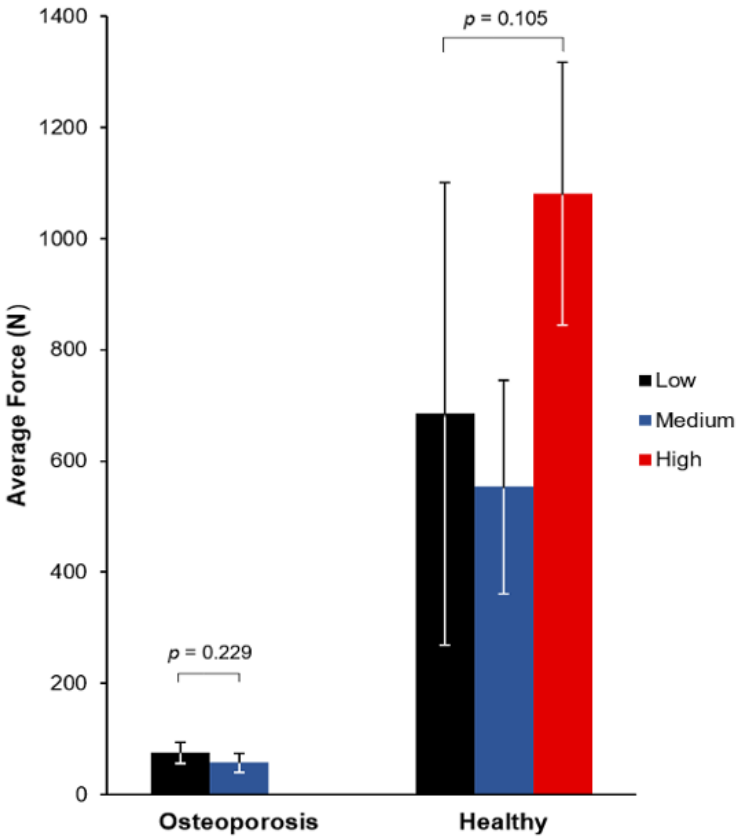


Figure 8-8 The average force required to remove the implant from osteoporotic and healthy femoral Sawbones® following impaction at different energy levels at 95% confidence interval
***p<0.05; **p<0.01; ***<p<0.001**

8.4. Discussion

8.4.1. Summary of results

This is the first biomechanical study to use DIC technology to analyse strain in synthetic femurs and highlights the importance of impaction technique for THA. As hypothesised, increased impaction velocity led to higher levels of peak strain (fracture risk) but did not significantly affect implant stability. In other words, the higher the impaction energy, the higher the resultant strain and the greater risk of periprosthetic fracture. In terms of implant stability, there was no statistically significant difference among the different velocities in both normal and low BMD samples, meaning that stability remains the same regardless of the amount of force.

The reason for this phenomenon may be explained by works by Doyle *et al.* and Mold *et al.*, who both highlighted a theory in their works known as the "law of diminishing returns" whereby increased impaction velocity fails to result in increased implant stability and furthermore, that a limit is reached beyond which any additional energy produces diminishing additional fixation (444, 455). This is highlighted in Figure 8-8, which shows that higher impaction does not result in better implant stability as represented by pull-out force.

For both specimen types, a strain behaviour pattern was identified in the femur during broaching. During impaction, a compressive peak strain was subsequently followed by a tensile peak strain, as highlighted in Figure 4c. Both mean peak strain and mean residual strains across all Gruen zones (GZs) were tensile in the direction of highest deformation. Following the peak tensile bone strain, a period of oscillation occurred which decayed and resulted in post-strike tensile strain. Despite no fractures occurring during mechanical

testing, both peak and tensile compressive strains are potential contributors to femoral fractures. A similar pattern of strain has been highlighted in a previous study (444), in which peak tensile strain, peak compressive strain and a similar period of oscillation resulting in a post-strike strain in synthetic acetabular cups can be identified.

In the healthy and osteoporotic bone specimens, the highest tensile strain levels – peak and residual, were greatest on the lateral aspect of the femur (GZ 1-3) in the direction of highest deformation, for all impaction velocities. Previous studies have demonstrated that the risk of fracture is increased when bones are in a tensile state, as a result of weakening of the cortical bone, indicating that the higher the tensile strain, the greater the fracture risk (435, 456-458). Thus, applying our results with those in the literature it could be suggested that if fractures had occurred in this study, they would most likely have occurred along the lateral femoral axis, equivalent to GZ 1-3. One hypothesis previously reported that the fracture pattern could be ascribed to both the configuration of the femoral rasp and the femur thus initiating implant-bone contact stresses (459). Furthermore, a study by Cummins *et al.* found that although not reaching statical significance, there was a correlation between the GZ where the largest strain was recorded and the location of the fracture (433). This was further supported by Flannery *et al.*, despite most of the fractures occurring on the medial zone of the proximal femur, their study found that the location of maximal strain recorded was closest to that of the fracture, with the number of specimen recording maximal strain at failure in lateral zones half the number of those in the medial zones (n=10/ n=19) (434). Multiple factors could account for the variation in the location of maximum strain. In their study, the researchers employed a model that utilised femoral morphology of revision hip arthroplasty; they used long stem Exeter implants cemented into femurs fixed in cement and applied an

Instron machine of increasing force. On the other hand, in our study, we utilised synthetic bone models to represent both healthy and osteoporotic bone conditions. These models were prepared and broached for primary short stem cementless femoral implant using a validated rig that accounts for soft tissue compliance.

8.4.2. Strengths and limitations

This work should be considered in the context of several limitations. First of all, synthetic bone does not require preservation and replicates similar mechanical properties and minimises the variability among specimens (460, 461). However, the differences in physical properties are likely to result in differences in the mechanical response during impaction, with the artificial bone geometry only offering an estimation of the complex pelvic anatomy and in vivo operative fixation. Future cadaveric studies are needed to validate our preliminary findings. Furthermore, the number and location of surface points used for the digital image correlation (DIC) and the accompanying GOM software varied among the samples and were influenced by several variables. These factors included the limited coverage of surface component, as only two cameras were used for capturing images, and the influence of bone curvature. Furthermore, the characteristics of the artificial stochastic pattern may play a role in determining the selection of surface points. The use of multicamera stereo-DIC technology presents a promising opportunity to capture three-dimensional views of specimens (462). Likewise, strain measurements in our study targeted the proximal femur to just below the femoral stem, failing to take into account strain behaviour related to Vancouver type C fractures (see Section 1.5.4), which represents approximately 24.6% of femoral periprosthetic fractures in cementless THR (410). Moreover, our impaction protocol employed a vertical loading axis with a 7° tilt. Whilst this approach yielded a standard testing environment across all specimens, it is

important to acknowledge that altering the inclination angle is poised to engender a spectrum of strain behaviours and fracture patterns, thus, future studies ought to investigate the effect of the loading angle on bone strain behaviour. Lastly, while our validated impaction rig provides a controlled and reliable method for research purposes, it is important to consider that surgical impaction in operative procedures tends to be less precise and accurate, which may potentially impact the significance and applicability of our data in practical settings.

Primary implant fixation is a critical factor in determining the technical success of the arthroplasty and patient morbidity. Based on this study, recommendations for clinical practice cannot currently be made, however, the study currently suggests that multiple lower impaction strikes may be preferable for the cementless technique to minimise fracture risk and maximise implant stability.

8.4.3. Conclusions

The results of this study may suggest that in normal- and low-density bones, a high number of low-energy strikes during femoral broaching and implant seating may provide lower peak strains with similar or increased implant stability. Presently, recommendations for clinical practice cannot be made from this study due to it being conducted on Sawbones®. The next chapter uses cadaveric specimen to validate these findings and assess the optimum impaction technique for normal BMD bones. Future studies should consider other variables such as strike orientation, implant displacement, and mallet mass to help establish an optimal cementless impaction technique.

Chapter 9 A cadaveric study on the effect of impaction force on cementless femoral periprosthetic fracture risk and implant stability in normal quality bone

9.1. Introduction

The previous chapter demonstrated that in normal- and low-density synthetic bones, higher impaction energies during cementless short-stem femoral broaching and implant seating may lead to higher fracture risks but similar implant fixation outcomes irrespective of the energy level. Whilst these artificial skeletal structures mimic the anatomical and mechanical properties of a human femur, their efficacy must be rigorously validated through cadaveric studies to adjust for the biomechanical complexities of human bones. To this regard, the utilisation of cadavers of normal bone mineral density (BMD) was employed in this study to delineate the biomechanical features associated with different impaction energy levels.

This study, involving femoral short stem cementless fixation in normal BMD cadavers, pursues a threefold set of objectives:

1. To compare the effects of three different energy levels on implant seating and femoral periprosthetic fracture risk during femoral broaching and implant seating,
2. To quantify and compare principal and residual strains in four Gruen zones for different energy levels, and
3. To illustrate femoral periprosthetic fracture threshold pattern exhibited upon impact to failure per energy level.

9.2. Materials and Methods

The investigations conformed with the research principles related to the Human Tissue Act (2004), and the study protocol was approved by the ethics board and Imperial College Healthcare Tissue Bank (reference number: R14088-3A, HTA licence 12275; Appendix 1, A1.2). The testing took place in a dedicated biomechanics laboratory at Imperial College London.

9.2.1. Specimen preparation

Sixteen fresh-frozen cadaveric hips (5 male; age, 48 ± 9 years; body mass index [BMI], 25 ± 5) were acquired from a tissue bank. The specimens had another use to examine young adult hip pathologies. A computed tomography (CT) scanner (Somatom Perspective; Siemens, Erlangen, Germany) was used to screen donors. Although some of the specimens indicated morphologies of femoroacetabular impingement (cam or pincer / acetabular retroversion), or developmental dysplasia of the hip, none of the specimens indicated any other hip pathologies (such as slipped capital femoral epiphysis and Perthes disease); arthritis; joint-space narrowing (< 2mm); or had a documented history of neuro-/musculo-skeletal trauma, injury, surgery, disability, or cancer. Imaging data were used to quantify bone morphology, particularly neck angle and femoral anteversion (Table 9-1).

Parameter	Measurement
Specimens (n)	16
Sides (left:right, n)	8:8
Mean age, years (SD)	48 (9)
Mean body mass index, kg/m^2 (SD)	25 (5)
Mean femoral head diameter, mm (SD)	44 (4)
Mean femoral anteversion, ° (SD)	15 (6)
Mean femoral neck-shaft angle, ° (SD)	129 (3)
Mean acetabular version, ° (SD)	25 (8)
Mean acetabular centre-edge angle, ° (SD)	28 (4)

Table 9-1 Descriptive demographic and anatomical parameters of the specimens

Cadavers were denuded to bone and frozen. Prior to testing, each specimen was defrosted and the distal end of femurs mounted in metal fixtures using methyl methacrylate bone cement and distal screws drilled into the cement to achieve a stable metal-cement construct. All specimens were prepared using a standard THA direct anterior approach and instruments for cementless THA (Furlong Evolution, JRI Orthopaedics, Sheffield) (442). Bone preparation is described in the previous chapter (see Section 8.2.1) and included: femoral neck resection; accessing the medullary canal with a box osteotome followed by a T-handled taper reamer in the posterolateral corner aligned with the femoral shaft axis; and the removal of medullary canal cancellous bone using the smallest broach size (size 6).

9.2.2. Strain measurement

Unlike synthetic bones, creating an adhesive contrasted stochastic speckle pattern on biological specimens is challenging due to their sleek lubricating physiological texture; paint film delamination and dissolution under high tensile deformation; and possible alterations to the native tissue mechanics, all of which are known to significantly affect the

DIC measurements (463). Given the advancements in materials science and speckle pattern fixation (440, 451, 463-465), we pilot-tested the application of speckle patterns to three distal femur cadaveric specimens, as described by Hensley *et al.* (466). In one specimen, we trialled the use of an abrasive paper for surface refinement, which may affect biomechanical properties as a trade-off when opting for DIC as the chosen technique. Despite this, we were unable to gain adhesive and durable speckle patterns suitable for DIC testing.

Electro-based mechanical strain measurement, such as strain gauge techniques, convert electrical signal changes into strain values and are commonly used as the 'gold standard' approach in the measurement of strain in biological tissues (467). In contrast to optical full-field three-dimensional DIC techniques, single strain gauges measure strain in a single direction and at the application site. Strain gauge rosettes, on the other hand, are composed of two or more strain gauges oriented in three arbitrary directions and allow strain analysis in a biaxial state. In cadaveric femurs, strain gauge rosettes and DIC showed a correlation in measurements with a significant average difference of 74 ± 95 με in maximum principal strain and 52± 31 με in minimum principal strain (466). Taken together and for pragmatic reasons, we employed rosettes for strain measurements in this study.

9.2.3. Strain Gauge Rosette attachment

Four pre-wired three-grid rosettes (350 Ω; 3 mm; 3-leads; wire length 3 metres, KFH-3-350-D17-11L3M3S, Omega Engineering Ltd., Manchester, United Kingdom) (Figure 9-1) were bonded to the anterior aspect of each proximal femoral cortex at identical positions in three different co-planar angles (0°, 45°, 90°), allowing for the calculation of first and second principal strains (ε1, ε2) and their corresponding angles.

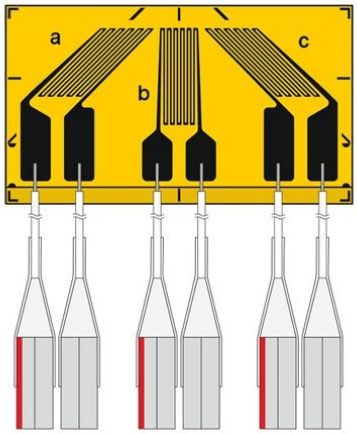


Figure 9-1 Pre-wired Rosettes oriented in three co-planar angles (0°,45°,90°)
Copyright: Omega Engineering Ltd

Positions for the rosettes followed a previously described standardised procedure (468-470). The two proximal rosettes were placed at the level of lesser trochanter (M1) and at the same level just below the greater trochanter laterally (L1), corresponding to Gruen zones 7 and 1, respectively. The two other rosettes were placed 32 mm below the lesser trochanter medially (M2) and laterally (L2), corresponding to Gruen zones 6 and 2, respectively (Figure 9-2). Prior to mounting the strain gauges, the cortical surfaces were further dissected removing remnants of periosteum, refined using delicately textured abrasive paper, and then degreased with acetone (471-474). All strain gauge rosettes were mounted parallel to the vertical femoral shaft axis using industrial grade superglue GP (CYN20; viscosity at 25° - 100 cps; set time range 5-15 seconds, Everbuild, Leeds, United Kingdom) followed by the application of cling film for adequate mounting (471-474).

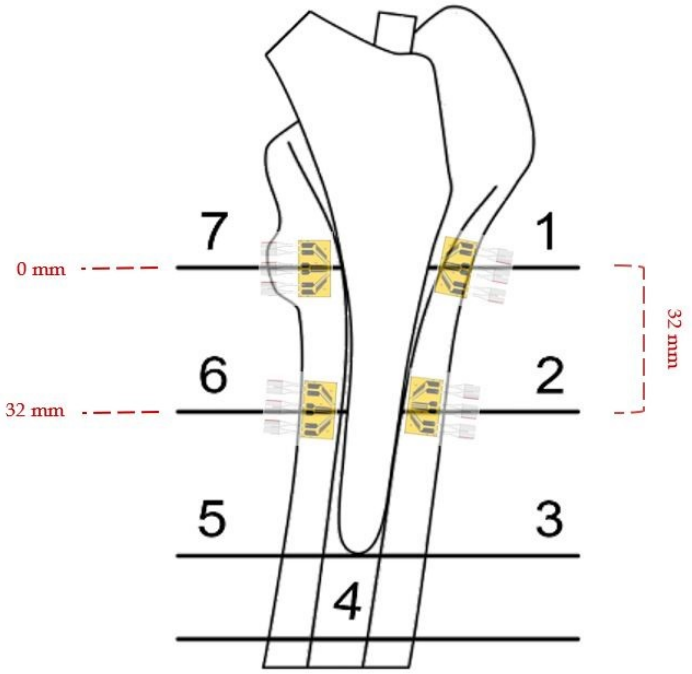


Figure 9-2 Positions of strain gauge rosette on the proximal femur

9.2.4. Set-up

When a specimen was ready for testing, the distal cemented part was placed into the metal holder and secured onto the impaction platen (Figure 9-3) in a vertical loading axis. Strain gauge wires were then connected to the half-bridge completion circuitry, preamplifier, 24-bit analogue to digital converters and serial data processor. Data capturing and storage was done using a computer and MATLAB software (2015a, MathWorks, Massachusetts, USA).

Figure 9-3 Experimental setup using a validated in-vitro rig

A validated in-vitro rig (Figure 8-2), with spring-dashpot support to account for soft tissue compliance, was used for testing (see Chapter 8) (447). Similar to the previous chapter, a platen drop weight of 1.2 kg was chosen to mirror the typical mass of a surgical mallet used in total hip arthroplasty. Impacts were administered to each sample through the controlled descent of a dropping platen onto a securely positioned broach/implant handle beneath. Given that energy (J) is the product of the mass (platen drop, kg) and the square of the velocity (see equation below), three energy levels mimicking surgical practice were generated by releasing the platen from three pre-determined heights (m/s) whilst using a standard "mallet" mass of 1.20 kg (Table 9-2) (448).

$$KE\ (J) = \frac{m\ (kg) * v^2(m/s)}{2}$$

Where:

KE = Kinematic energy (joule); m = mass weight (kg); v^2 = velocity (m/s).

Energy Level	Low	Medium	High
Drop Mass (kg)	1.20	1.20	1.20
Drop Velocity (m/s)	1.45	2.84	4.08
Energy (J)	1.26	4.84	9.99

Table 9-2 Impaction energy levels

9.2.5. Testing

Sixteen fresh-frozen cadaveric hips representing healthy BMD were tested at low (n = 5), medium (n = 5), and high (n = 6) velocities.

9.2.5.1. Broaching and implant seating

Femoral broaches of increasing sizes ranging between seven and eleven, were impacted into the femoral canal. Three consecutive strikes were administered with the broach removed after each third strike, replicating the approach commonly seen in surgical procedures. The broach was deemed seated once the cutting teeth were parallel to the depth of the neck resection. The choice of implant size and its fixation relied on clinical assessment; however, the overarching strategy followed was to opt for a maximal implant size that the bone could safely accommodate.

9.2.5.2. Implant stability

After seating an implant, the twelve circuits were disengaged from the analogue and the metal platen and its specimen were attached to a uniaxial loading testing machine (Instron model 5565; Instron, High Wycombe, UK). Pull-out fixation tests evaluated implant

stability using an eleven-millimetre steel rod to connect the implant to the Instron's actuator head and then applying a tensile load at a displacement rate of 0.2mm/min (see Figure 8-4) (454).

9.2.5.3. Fracture threshold

Once the implant was removed, the platen and specimen were set-up again on the rig for the impaction-to-fracture phase. Using the same implant size, twenty consecutive strikes at each weight level were delivered, with increasing drop mass weight, ranging from 1.2 kg to 8 kg (Table 9-3). The test concluded either upon bone fracture or upon completing twenty impacts of the 8 kg drop weight.

Energy levels		**Low**	**Medium**	**High**
Drop Velocity (m/s)		1.45	2.84	4.08
			Energy (J)	
Drop Mass (kg)	**1.20**	1.2615	4.83936	9.98784
	1.50	1.576875	6.0492	12.4848
	1.80	1.89225	7.25904	14.98176
	2.10	2.207625	8.46888	17.47872
	2.40	2.523	9.67872	19.97568
	2.70	2.838375	10.88856	22.47264
	3.00	3.15375	12.0984	24.9696
	3.30	3.469125	13.30824	27.46656
	3.60	3.7845	14.51808	29.96352
	3.90	4.099875	15.72792	32.46048
	5.00	5.25625	20.164	41.616
	6.00	6.3075	24.1968	49.9392
	8.00	8.41	32.2624	66.5856

Table 9-3 Impaction-to-fracture energy levels per velocity

9.2.6. Data processing and measurements

Each single impaction generated an excel file of 50,000 rows containing: velocity; strike load; base load; and strain measurements over time for the twelve channels. The files per specimen were renamed and uploaded to MATLAB software (2021b, MathWorks, Massachusetts, USA). Adjustments related to the offset between post-strike strain and consecutive pre-strike initial strain were accounted for. A Butterworth electronic signal processing filter was applied to minimise high-frequency noise or fluctuations in raw data whilst enhancing the integrity and quality of strain measurements via providing a smooth transition between the passband and stopband.

First and second principal strains, as measures of deformation, together with their directions were calculated using the micro-strain (με) values and the following equations (Omega Engineering Ltd., Manchester, United Kingdom):

$$\boldsymbol{\varepsilon_{p,q}} = \frac{\mathbf{1}}{\mathbf{2}}\left[\boldsymbol{\varepsilon_1} + \boldsymbol{\varepsilon_3} \pm \sqrt{(\boldsymbol{\varepsilon_1} - \boldsymbol{\varepsilon_3})^2 + (\mathbf{2}\boldsymbol{\varepsilon_2} - \boldsymbol{\varepsilon_1} - \boldsymbol{\varepsilon_3})^2}\right]$$

$$\boldsymbol{\theta_{p,q}} = \frac{\mathbf{1}}{\mathbf{2}}\,\boldsymbol{TAN^{-1}}\,\frac{\mathbf{2}\,\boldsymbol{\varepsilon_2} - \boldsymbol{\varepsilon_1} - \boldsymbol{\varepsilon_3}}{\boldsymbol{\varepsilon_1} - \boldsymbol{\varepsilon_3}}$$

Where:

$\boldsymbol{\varepsilon_1}$ = strain in gauge 1; $\boldsymbol{\varepsilon_2}$ = strain in gauge 2; $\boldsymbol{\varepsilon_3}$ = strain in gauge 3; $\boldsymbol{\varepsilon_{p,q}}$ = principal strains; $\boldsymbol{\theta_{p,q}}$ = the acute angle from the axis of gauge 1 to the nearest principal axis. When positive, the direction is the same as that of the gauge numbering and, when negative, opposite.

First principal strain represents the magnitude of strain experienced by the bone and commonly refers to the tensile force corresponding to bone strength and fracture risk. Second principal strain refers to the shear loading, or compression force, experienced by the bone in a direction that is orthogonal to the first principal strain and is of significant

value to assess bone stability under shear forces. Maximal first and second principal strains were calculated and averaged inter- and intra-specimens.

Residual strain is the change in bone innate strain due to impaction, and provides information related to change in bone behaviour and material properties. The difference between post-strike residual and pre-strike baseline strains was measured at identical timepoints for each strike.

$$\varepsilon_R = \varepsilon_{post-strike} - \varepsilon_{pre-strike}$$

Where:

ε_R = residual strain per strike; $\varepsilon_{post\text{-}strike}$ = strain value at x axis value 49,000; $\varepsilon_{pre\text{-}strike}$ = strain value at x axis value 500.

With regards to the fracture threshold calculation, the energy delivered to each specimen per energy level over the entirety of procedural phases was calculated using the following equations:

$$E_{procedural\ step} = KE_{step} * n_{step}$$

$$E_{specimen}(J) = E_{B7} + E_{B8} + E_{B9} + E_{B10} + E_{Implant} + E_{1.2kg} + E_{1.5kg} \ldots$$

$$E_{total\ per\ energy\ level}(J) = mean\ (E_{specimen\ 1} + E_{specimen\ 2} \ldots)$$

Where:

KE_{step} = Kinematic energy (J) per procedural step; n = number of strikes delivered per step.

9.2.7. Statistical analysis

Strain variations resulting from strikes of the three velocities (low, medium, and high) were evaluated using a two-way T test for inter-group comparisons. A one-way Analysis of Variance (ANOVA) test, followed by Tukey's Honestly Significant Difference (HSD) post hoc test, was employed to assess differences among the groups. The level of

significance was established as $\alpha = 0.05$ for all tests. Strain data per group were normally distributed. Descriptive statistics are also presented. MATLAB software (2021b, MathWorks, Massachusetts, USA) was used to analyse all data.

9.3. Results

All specimens for high (n = 6), medium (n = 5), and low (n = 5) were tested for broaching, implant seating, and pull-out tests. Due to testing artifacts including dislodging of rosette strain gauges ($n_{low} = 2$) and testing error during pull-out tests ($n_{high} = 1$), these samples were not tested for the impaction-to-fracture steps and the latter was excluded from the pull-out test analysis. Two samples ($n_{low} = 1$, $n_{medium} = 1$) were excluded from the analysis due to dislodgement of rosette strain gauges and poor invalid data quality. Table 9-4 highlights the number of samples tested and analysed for each step.

	Tested			**Analysed**		
Procedural phase	**Low (n)**	**Medium (n)**	**High (n)**	**Low (n)**	**Medium (n)**	**High (n)**
Broaching and implant seating	5	5	6	4	4	6
Pull-out tests	5	5	6	5	5	5
Impaction-to-fracture	3	5	5	3	5	5

Table 9-4 Number of specimens tested and analysed per procedural steps

A greater number of strikes was required for the low velocity versus either medium/high velocities for broaching (one-way ANOVA; $p < 0.05$) and seating an implant (one-way ANOVA; $p=0.0458$), as shown in Table 9-5. No difference was detected regarding the number of strikes between medium versus high velocities for broaching (two-way t-test, $p > 0.05$ for different broach sizes) and implant seating (two-way t-test, $p = 0.47$). Figure

9-4 shows the mean number of strikes during mechanical testing at each impaction velocity.

	One-way ANOVA	Two-way t-test	
	P value	**Comparison groups**	**P value**
Broach 7	0.0002	Low vs medium	0.0022
		Low vs high	0.0016
		Medium vs high	0.2611
Broach 8	0.0083	Low vs medium	0.0022
		Low vs high	0.0792
		Medium vs high	0.1200
Broach 9	0.0007	Low vs medium	0.0022
		Low vs high	0.0069
		Medium vs high	0.1166
Broach 10	0.0256	Low vs medium	0.0161
		Low vs high	0.2113
		Medium vs high	0.0591
Implant	0.0458	Low vs medium	0.1110
		Low vs high	0.0484
		Medium vs high	0.4761

Table 9-5 Number of strikes among and between groups comparisons for different procedural steps per energy level

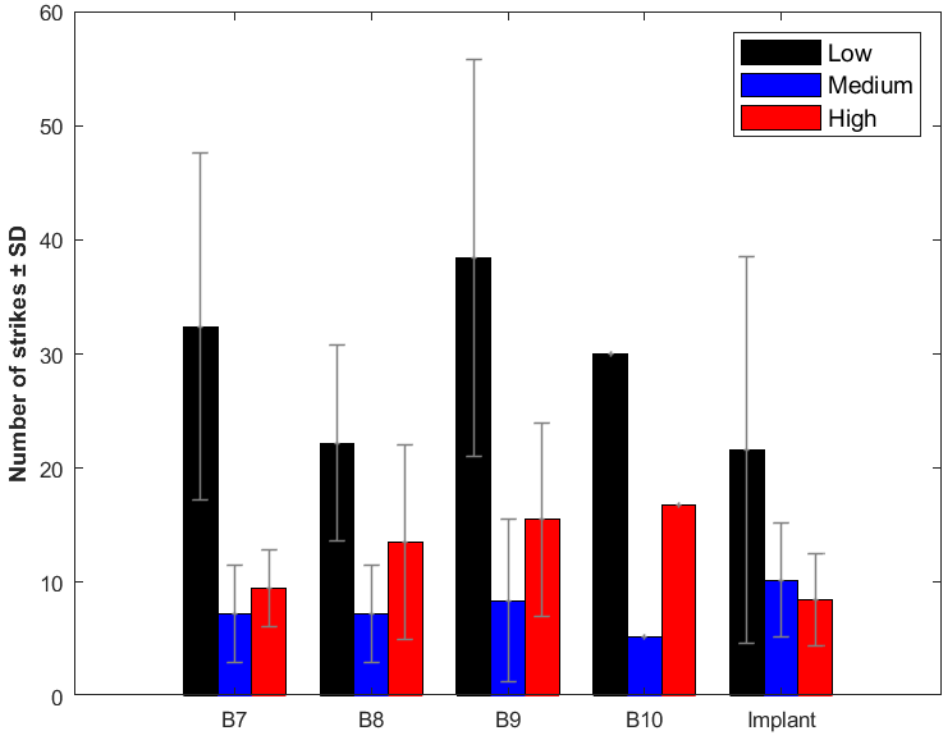


Figure 9-4 Average number of strikes required to seat broaches 7-10 and implant for different energy levels. The x-axis demonstrates the procedural step related to the size of broach used or implant seating

The average first principal strain was statistically significant among the three energy levels in the superior Gruen zones (one-way ANOVA; medial $p = 0.016$, lateral $p = 0.002$) (Figure 9-5). Two-way t-test comparisons, and post hoc Tukey's Honestly Significant Difference (HSD) adjustments, showed increased average first principal strain between the high and the two other energy levels in Gruen zone 7 (high-to-low $p = 0.045$; high-to-medium $p = 0.037$ but not low-to-medium $p = 0.189$) whilst in Gruen zone 1, only high-to-low was statistically significant $p = 0.004$. For the inferior region the average first principal strain increased in the inferomedial region (Gruen zone 6, one-way ANOVA $p = 0.0001$) and was only statistically significant when comparing high to low and medium velocities ($p \leq 0.001$ each) but did not differ between low and medium energies ($p = 0.99$). There was no statistically significant difference among the three energy levels in the inferolateral area (Gruen zone 2, one-way ANOVA $p = 0.418$) (Figure 9-5).

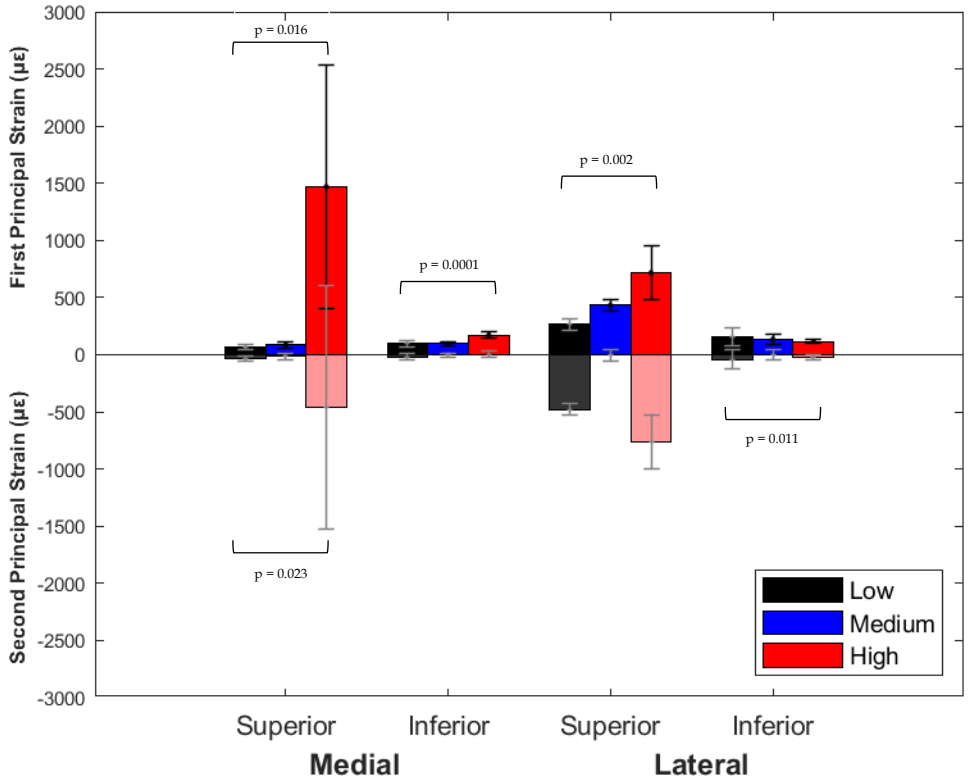


Figure 9-5 Average maximal first and second principal strains during broaching and implant seating per three velocities and four Gruen zones. The x-axis demonstrates the rosettes strain gauge locations on the cortical bone corresponding to the four Gruen zones: superomedial (GZ 7); inferomedial (GZ 6); superolateral (GZ 1); and inferomedial (GZ2)

The average second principal strain was statistically significant among the three energy levels in Gruen zones 2 and 7 (one-way ANOVA; p = 0.011 and p = 0.023, respectively). However, following inter-group comparisons the difference remained statistically significant only for the medium-to-high group in Gruen zone 7 (two-way t-test; p = 0.041, post hoc HSD; p = 0.046), whilst in Gruen zone 2 medium velocity was different to the other two groups (medium-to-low p = 0.003; medium-to-high p = 0.025).

With regards to the directions of first and second principal strains, they averaged between 20° and 30° and there was only a statistically significant difference in the superolateral region (Gruen zone 1, one-way ANOVA; $p_{\theta 1st}$ = 0.012, $p_{\theta 2nd}$ = 0.003) (Figure 9-6). Sub-group analyses, using post hoc HSD and a two-way t-test, in that zone showed no difference in the average direction of the first principal strain between low and medium

velocities but a statistically significant difference between high and low energies (two-way t-test: $p_{\theta 1st} = 0.029$; post hoc HSD: $p_{\theta 1st} = 0.045$) and between high and medium velocities (two-way t-test: $p_{\theta 1st} = 0.025$; post hoc HSD: $p_{\theta 1st} = 0.028$), whilst the average direction of the second principal strain was only statistically significant between the medium and high groups (two-way t-test: $p_{\theta 2nd} = 0.004$; post hoc HSD: $p_{\theta 2nd} = 0.003$).

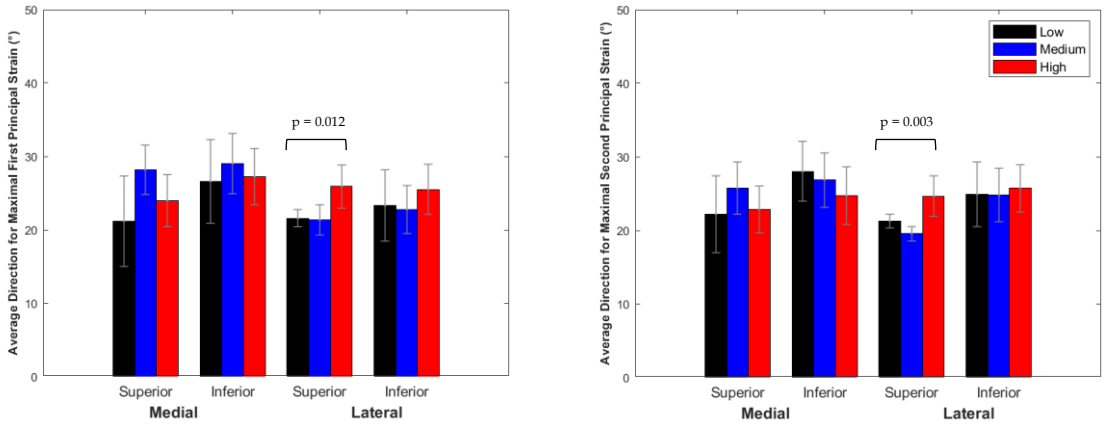


Figure 9-6 Average direction for maximal first and second principal strains during broaching and implant seating per three velocities and four Gruen zones. The x-axis demonstrates the rosettes strain gauge locations on the cortical bone corresponding to the four Gruen zones: superomedial (GZ 7); inferomedial (GZ 6); superolateral (GZ 1); and inferomedial (GZ2)

In terms of average cumulative residual strain, similar to the first principal strain, high velocity yielded the highest residual strain in three of the Gruen zones (1, 6 and 7). There was a statistically significant difference among groups in these zones (one-way ANOVA: superomedial $p = 0.002$; inferomedial $p = 0.011$, superolateral $p = 0.001$) but not in the inferolateral region ($p = 0.388$) (Figure 9-7). In the superior Gruen zones, there was an increased average residual strain between high and the two other energy levels in Gruen zone 7 (two-way t-test: high-to-low $p = 0.005$; high-to-medium $p = 0.026$, medium-to-low $p = 0.002$, post hoc HSD: $p < 0.001$ each) and Gruen zone 1 (two-way t-test: high-to-low $p = 0.006$; high-to-medium $p = 0.013$, medium-to-low $p = 0.234$, post hoc HSD: $p < 0.001$

each). Sub-group analyses for the inferomedial region (Gruen zone 6) showed a statistically significant difference between high velocity and the low (two-way t-test p = 0.040, post hoc HSD p = 0.049) and medium (two-way t-test p = 0.019, post hoc HSD p = 0.021) but not between low-to-medium velocities (two-way t-test p = 0.621, post hoc HSD p = 0.971).

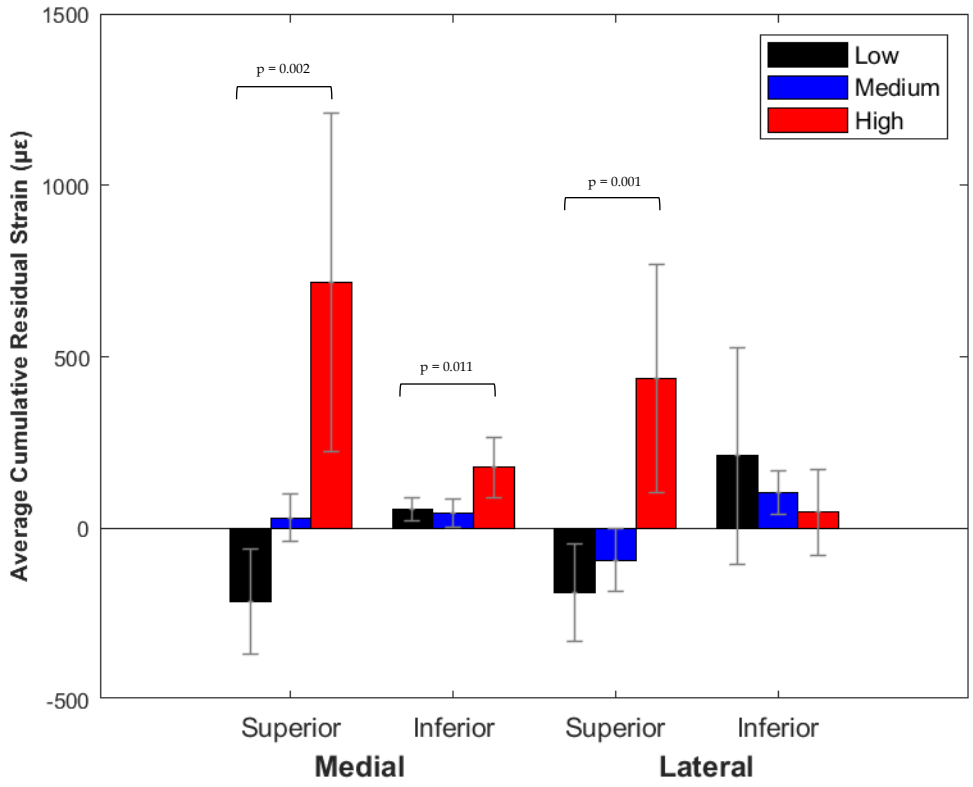


Figure 9-7 Average cumulative residual strain during broaching and implant seating per three velocities and four Gruen zones. The x-axis demonstrates the rosettes strain gauge locations on the cortical bone corresponding to the four Gruen zones: superomedial (GZ 7); inferomedial (GZ 6); superolateral (GZ 1); and inferomedial (GZ2)

There was no difference in the force required to remove specimens impacted by low, medium, or high velocity strikes (one-way ANOVA: p = 0.208; two-way t-test: high-to-low p = 0.773, high-to-medium p = 0.058, medium-to-low p = 0.200) with a mean of 1,605 N, 1,065 N and 1,731 N, respectively (Figure 9-8).

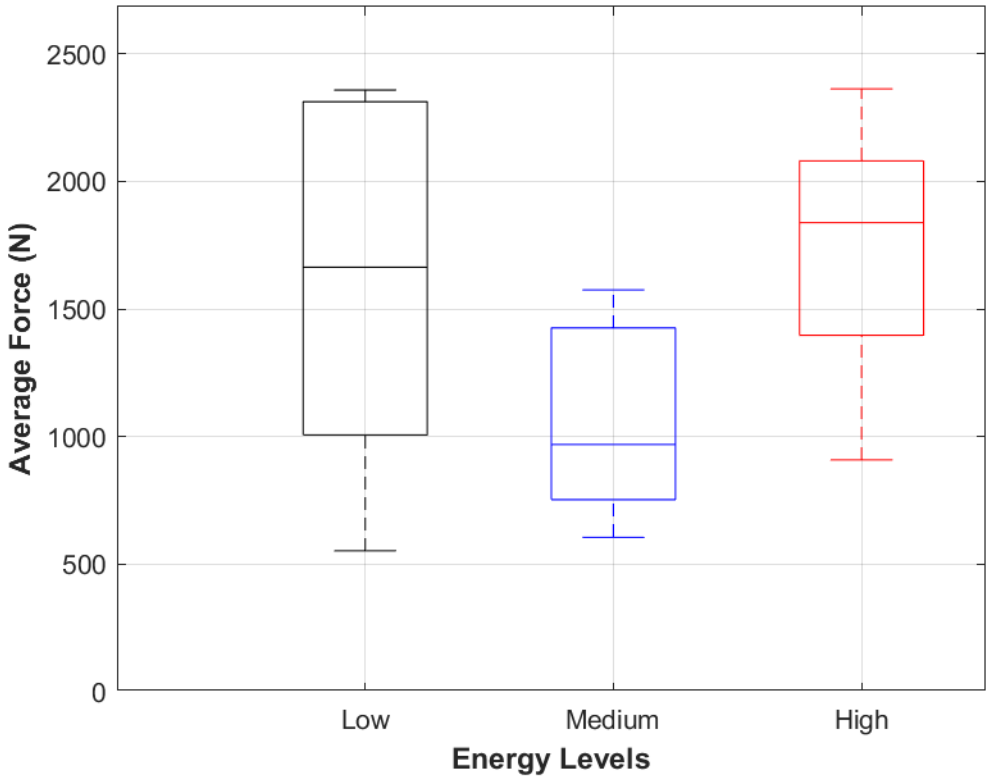


Figure 9-8 Average force during the pull-out test per three velocities and four Gruen zones (p = 0.208). The x-axis demonstrated the energy levels used: low, medium, or high

Figures 9-9, 9-10, and 9-11 highlight the changes in average maximal first, second principal, and residual micro-strains over the procedural phases at each Gruen zone. Overall, all high (n_{high} = 5) and medium (n_{medium} = 5) velocity specimens fractured along the procedural steps post implant seating and removal with the earliest fracture happening at the 1.2 kg step with the later fractures ending at the 2.1 kg and 2.4 kg steps, respectively. Two out of the three low energy samples (n_{low} = 2) tested for fracture threshold did not fracture despite delivering 260 strikes with increasing mass weight from 1.2 kg to 8 kg whilst one specimen (n_{low} = 1) fractured at the 6 kg step.

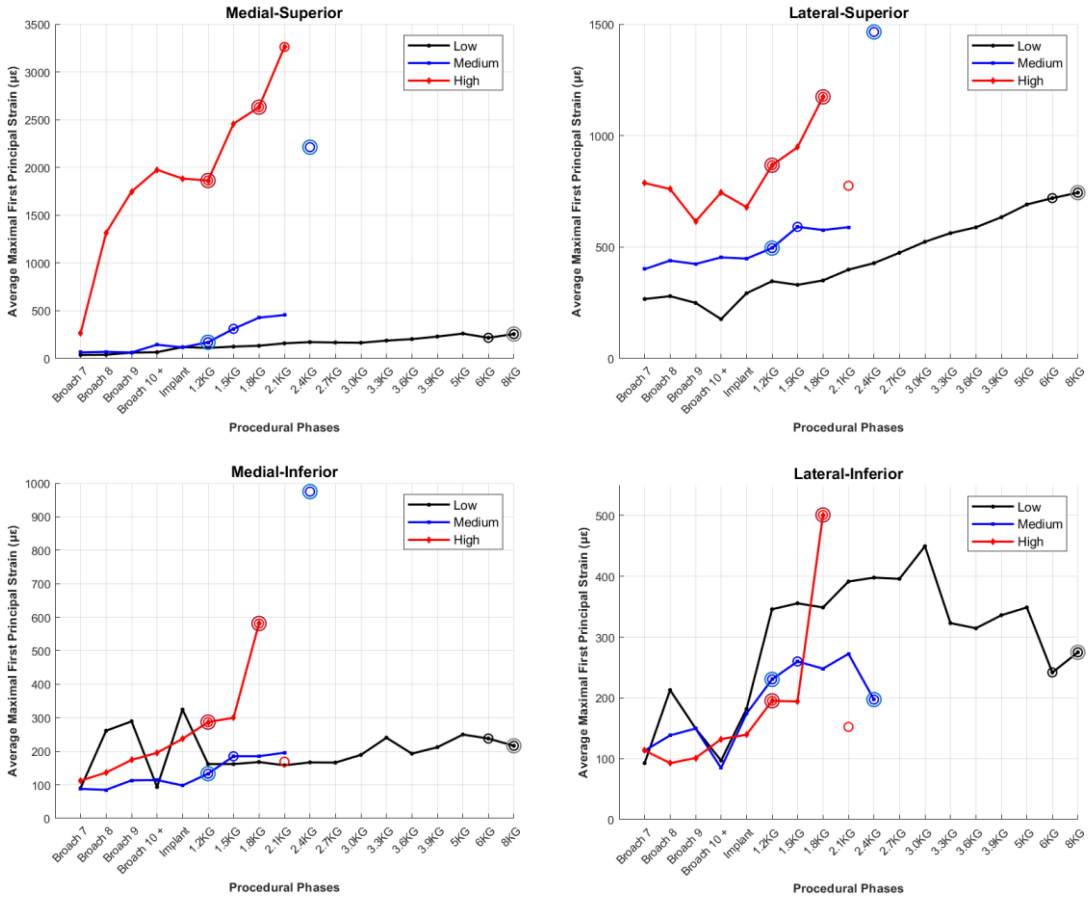


Figure 9-9 Average maximal first principal strain per procedural phases
the position and number of hollow circles indicate fracture occurrence

In the superior zones, first principal (Figure 9-9) and residual strains (Figure 9-11) for the high velocity specimens were significantly higher than the other two groups throughout the procedural steps and just before fracture occurrence in the medium energy group, whilst no striking differences were witnessed in the inferior zones.

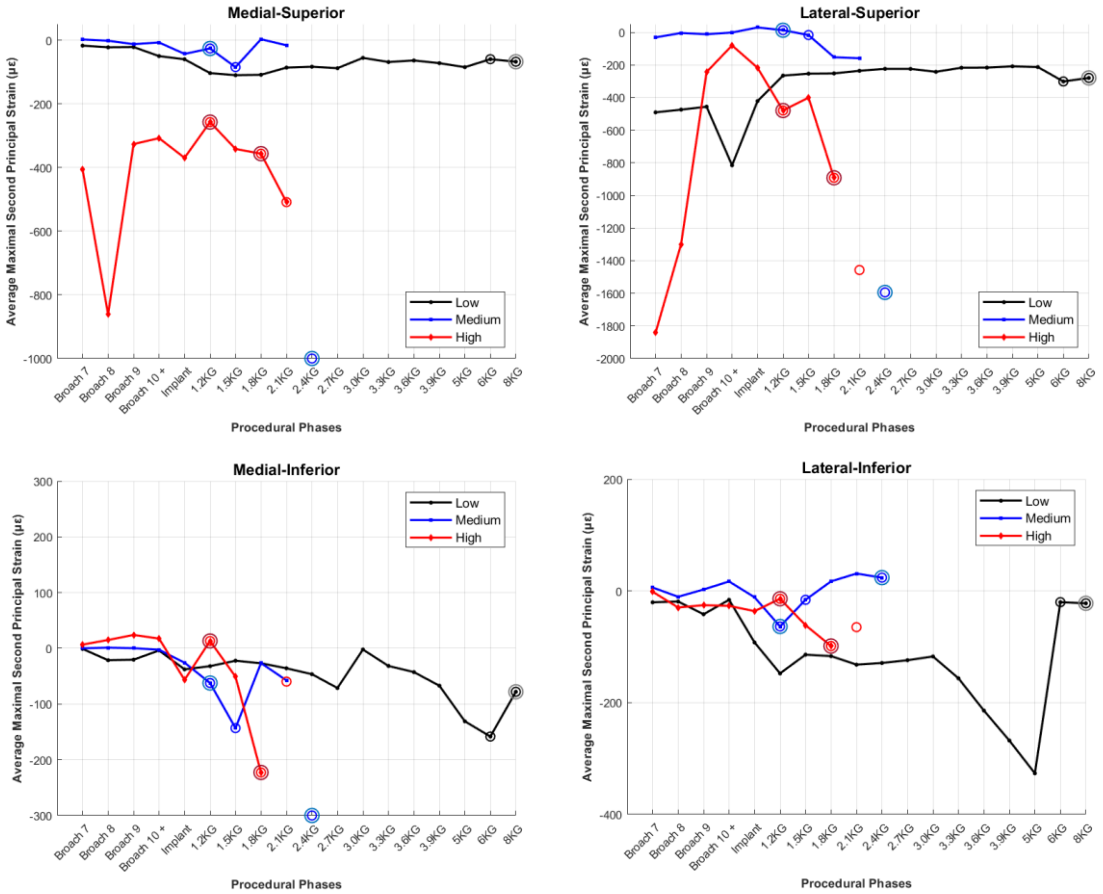


Figure 9-10 Average maximal second principal strain per procedural phases
the position and number of hollow circles indicate fracture occurrence

In the graphs pertaining to the second principal strain (Figure 9-10), high velocity samples appeared to have lower micro-strain values in the superior zones whilst no significant differences among the groups were observed in the inferior Gruen zones.

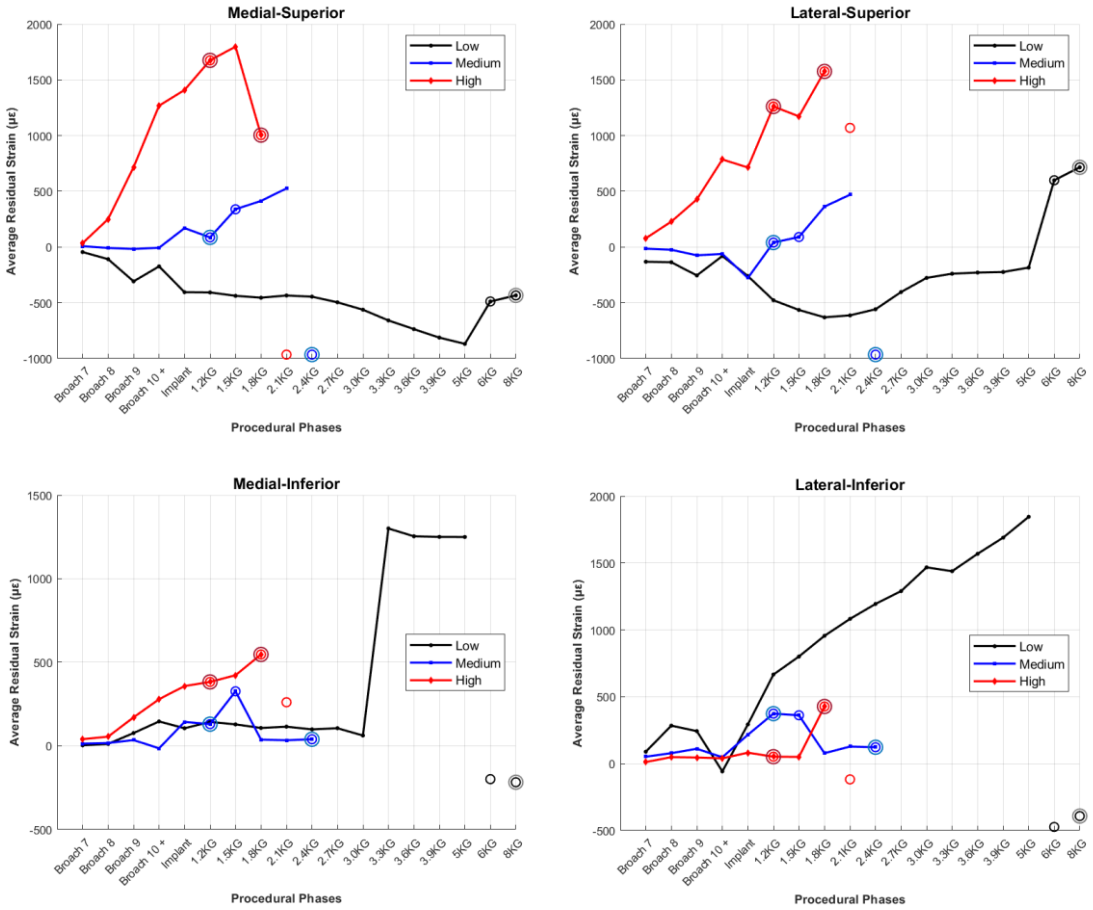


Figure 9-11 Average residual strain per procedural phases
the position and number of hollow circles indicate fracture occurrence

The total accumulated energy delivered for each specimen per energy group over the entire procedural phases are shown in Figure 9-12. Despite the earlier occurrence of fractures in the high velocity group (Figure 9-9), the accumulated energy delivered to the bone did not significantly differ among the three energy groups (one-way ANOVA p = 0.08, two-way t-test: high-to-medium p = 0.054; high-to-low p = 0.461, medium-to-low p = 0.082). The average fracture thresholds per each energy group were: high (n = 5, 1390.7 J [426 – 2600]); medium (n = 5, 496.0 J [145 – 953]); and low (n = 1, 893.3 J).

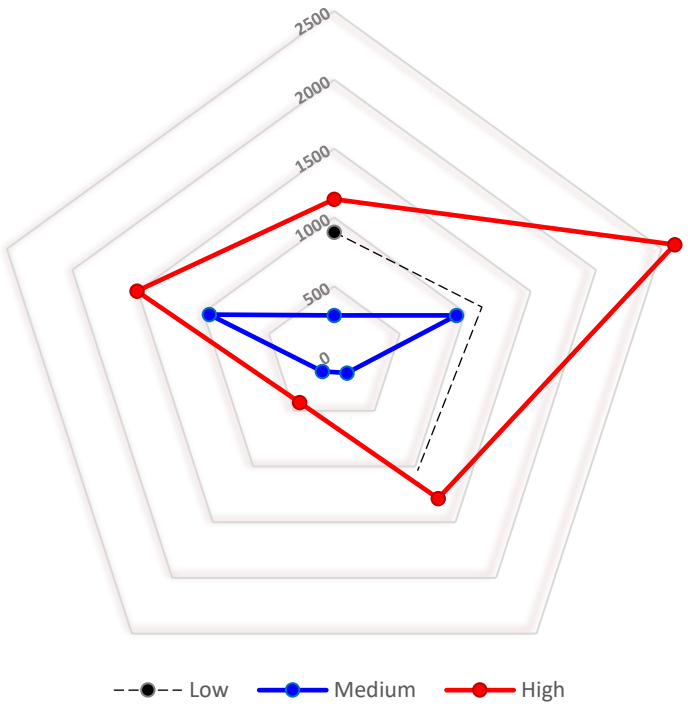


Figure 9-12 Accumulated energy (J), ranging from 0 – 2500, delivered to each specimen over the entirety of the procedural phases. Markers indicate fractures; all high- and medium-energy specimens fractured, whilst only one (n = 1 out of 3) low-energy specimen fractured. The dotted line indicates the accumulated low-energy pattern for the other two non-fractured hips

9.4. Discussion

9.4.1. Summary of results

Our findings closely resemble the outcomes observed in the preceding investigation on synthetic bones, demonstrating that higher energy impactions during cementless femoral broaching and implant seating in cadaveric specimens are associated with increased levels of first and second principal strains (periprosthetic fracture risk and bone deformation) without affecting implant fixation to the bone. In this study, we have also shown that whilst lower force significantly requires a higher number of strikes, the post-strike residual strain is augmented with high energy impactions which may also contribute to the early fracture pattern witnessed with higher energies.

The first and second principal strain levels corresponding to tensile (fracture risk) and compressive (shear loading or deformation) forces, respectively, were greatest in the proximal femur, Gruen zones 1 and 7, in the 23° - 28° direction. For low and medium velocities, tensile strain was highest on the superolateral side of the femur and this is in agreement with our previous study suggesting that fractures during impaction are more likely to occur due to increased tensile strain (435, 456-458) along the lateral femoral axis (Gruen zones 1 and 2), whilst compressive strain remained minimal and did not significantly differ between low and medium energies. However, principal strains for the high energy group were highest in both superior Gruen zones: tensile (medial > lateral) and compressive (lateral > medial).

A handful of studies investigated the effect of impaction on femoral strain behaviour. Flannery *et al.'s* experiment using an axial Instron loading machine for long-stem cemented impaction in revision hip arthroplasty recorded double the number of fractures in the medial zones (n = 19) compared to the lateral side (n = 10), which may explain the elevated first principal strains in the upper-medial Gruen zone for the high energy cohort in our findings (434). Nonetheless, whilst the authors reported wide variation in bone surface strain with failure ranging between 1,040 and 15,800 micro-strains, data was not available to correlate whether medial femoral axis fractures occurred at higher micro-strain levels. Despite the differences in their study design (see Section 8.4.1) and considering that fracture pattern could also be attributed to configuration of the femoral rasp (459), it can be argued that higher energy impactions may increase fracture risk in the superomedial and superolateral zones whilst low and medium energy levels increase bone tensile strain along the superolateral femoral region.

This phenomenon can be further explored through analysing cumulative residual strain in our study, which for the high energy group was greatest and tensile (positive) in the superior aspect of the femur (medial > lateral) with statistically significant differences compared to the compressive (negative) residual strains of medium/low velocities. This may indicate that during high energy impactions, medial compressive strain may transition to a tensile state under substantial energy loads. Recent studies investigating stress shielding through both in-vitro quasi-physiological experimental assays and finite element analyses (FEA), using both press-fit polyether ether-ketone hip and control generic cementless THR (Ti6Al4V) prostheses, demonstrated similar stress distribution of lateral tensile strain and medial compressive strain. However, the highest stress shielding increase was seen in both superomedial and superolateral Gruen zones, which support our findings (470, 475). Nevertheless, the researchers employed an axial loading model using the Instron machine of increasing force from 500 - 1200 N on synthetic bones to replicate daily activities rather than impaction, making them incapable of reproducing high-energy collision scenarios seen in femoral impaction.

In terms of implant stability, there was no statistically significant difference among the different velocities, meaning that implant stability remains the same regardless of the amount of force per strike. Likewise, our findings show that a greater number of strikes was required for both broaching and implant seating in the low velocity cohort compared to medium and high velocities. This is in agreement with a recent study by Depuy-Synthes analysing the parameters of press-fit femoral stem impaction and that found, whilst lacking experimental bone strain data, that a greater number of lower energy hits reduces both the axial stresses in the introducer and the manual strength required by surgeons without affecting implant fixation (476).

In relation to the cumulative energy administered during the procedural stages, it was observed that while the high-energy group received heightened energy levels through fewer strikes, no statistically significant distinction was evident among the groups. Similarly, two specimens in the low-energy group failed to fracture despite reaching comparable cumulative force to medium and high velocity. This implies that the incidence of fractures is likely attributed to the increased tensile force generated, rather than the cumulative energy levels in the bone. This is in agreement with previous literature demonstrating a correlation between bone tensile state and fracture risk as a result of cortical bone weakening (435, 456-458). No previous studies investigated fracture threshold during cementless femoral impaction in primary hip arthroplasty. Cummins *et al.* identified threshold force using Instron axial cyclic loading in cemented long-stem revision hip arthroplasty to be 0.5 kN with the majority of cadaveric femurs (15/17) fractured at or above 1.6 kN (433). Whilst our study measured the cumulative force at fracture point in healthy cadaveric femurs to be 893 J, 496 J, and 1390 J for low-, medium- and high-energy hits, respectively, this is of limited significance given the wide confidence intervals and with the low fractured sample size (n = 1) for the low-energy group, however, it may provide a foundation for future studies in this field.

9.4.2. Strengths and limitations

This work should be interpreted considering several limitations. The three limitations discussed in the preceding chapter (Section 8.4.2) persist and include:

1) The absence of strain measurements in the distal femur results in a significant omission of a quarter of Vancouver type C periprosthetic fractures occurring in cementless THR;

2) Strike orientation was not uniform across all specimens for pragmatic testing reasons, which may significantly alter strain behaviours and fracture patterns; and
3) The controlled testing environment might not accurately emulate the constrained precision in surgical practice.

In addition, limitations related to the use of strain gauge rosettes are well-chronicled in literature and were discussed in Section 9.2.2. Notably, strain gauges provide limited point strain measurements; are amenable to temperature effects; and misalignment and positioning issues resulting in inaccurate strain readings. To mitigate these challenges, our study followed a previously described standardised procedure with the rosettes placed at identical positions and alignments whilst conducting the experiments under exact testing conditions. Finally, the study utilised Furlong Evolution instruments (JRI Orthopaedics, Sheffield) and the outcomes may differ with alternative implant designs.

9.4.3. Conclusions

This study highlights that there is an increased probability of fracture risk and bone deformation in the proximal femur, particularly in the superomedial area, resulting from high-energy impacts during cementless femoral broaching and implant seating conversely, the incidence is considerably reduced with low or medium velocity strikes, while implant fixation remains uniform across varying energy levels. However, further investigations are warranted to understand the effects of competing variables such as strike orientation, implant design and rosettes' positioning - all of which can potentially influence the observed strain behaviour.

Chapter 10 Discussion

10.1. Summary of findings

The work presented in this PhD thesis has provided valuable understanding into the predictive factors and outcomes in primary elective total hip replacement. The dissemination of findings aspires towards a more informed and collaborative shared decision-making approach. Using personalised risk stratification models in surgical practice may help balance the safety, implant failure, and functional outcomes in the context of implant selection and patient characteristics. The work alluded to a heterogenous technique for cementless femoral broaching and implant seating, wherein non-clinical laboratory-based recommendations for surgical impaction were provided to minimise the risks of intra-operative periprosthetic femoral fractures and loosening.

10.1.1. Part I: Machine learning and outcomes following primary THR

The study outlined in Chapter 3 represents the largest, dual-validated registry-based machine learning study forecasting post-operative safety and failure outcomes following primary THR. ML models outperformed conventional logistic regression in predicting all outcome variables in both databases. Whilst forecasting mortality risks achieved "good" binary discriminative power with area under the curve (AUC) > 0.80, predicting implant failure risk was more challenging (SHAR $AUC_{revision}$ = 0.72, $AUC_{re\text{-}operation}$ = 0.77; NJR $AUC_{revision}$ = 0.61). Five studies developed machine learning models to predict the risks of mortality, revision, or re-revision following primary elective THR using different datasets from Norway, Sweden, the United Kingdom and the United States (377-381). Whilst their predictive power was comparably reported to range between 0.75 and 0.80 for mortality and 0.80 – 0.87 for implant failure, significant limitations variably hindered their

generalisability and applicability in clinical practice. These included the exclusion of uncemented implants, limited sample size, the absence of patient censoring in survival modelling, and under-representation of the studied outcomes. The Bland-Altman estimates reported in Chapter 3 highlighted no systematic difference in the AUC measurements between the two registries with a narrow line of bias suggesting a high level of concurrent validity of the developed models. Garland *et al.* reported building a web-based calculator for mortality risk, however, the user interface (UI) appears to be inaccessible (377, 477). Future studies should seek to implement and validate the algorithms in clinical settings with emphasis on intuitive UI and user experience (UX) as well as the principles of behavioural science to mitigate possible resistance, ethical considerations related to patient consent and data privacy, and maintenance strategies and reinforcement learning techniques to optimise predictions over time.

Chapter 4 identified the patient, surgical, and implant-specific factors associated with poor mortality and revision outcomes using a ML variable-selection approach. A statistically significant trade-off pattern was found encompassing a heightened mortality risk in patients with low body mass index (BMI) and an augmented revision risk in obese individuals. Trela-Larsen *et al.* implemented a flexible parametric survival regression to provide individualised estimations of one-year mortality risk after elective hip and knee arthroplasty using the NJR as a training platform and externally validated their model using the Norwegian Arthroplasty Register (NAR) (378). The authors reported in their supplementary materials a U-shaped curve for mortality hazard ratio for categorical, linear, and non-linear fractional polynomial functions of body mass index, indicating that underweight and obese patients were both at increased risk of one-year mortality following primary THR – a phenomenon described in Chapter 4. Furthermore, patients with higher co-morbidities or with a pre-operative diagnosis of non-traumatic avascular

necrosis were at significantly increased risks of unfavourable outcomes. Elderly patients were at a small increased risk of mortality but a lower revision risk, whilst males exhibited higher risks for both outcomes. Similar to these findings, previous ML models identified increasing age, male gender, higher co-morbidities, and body mass index to significantly increase mortality and revision risks in primary THR (377, 380). However, considering prior literature, age and gender were likely attributed to confounders and were adjusted for in the multivariate analyses. Of importance in Chapter 4, the fixation method stood out as a significant modifiable prognostic factor wherein cemented fixation was associated with higher mortality rates, but lower rates of implant failure compared to cementless fixation. Whilst this agrees with prior studies on the Swedish and Norwegian registries (348, 378), statistical significance for mortality risk was not found in the NJR (HR 1.06, 95% CI 0.96 to 1.17, p = 0.25). It was acknowledged, however, that the study in Chapter 4 did not evaluate the reasons for revision nor adjusted for various co-variates that could have augmented the effect of confounding by indication. These limitations informed Chapter 6 to predict the risk of periprosthetic femoral fracture necessitating re-do surgery, and Chapter 7 to investigate the effect of implant fixation on safety and implant failure outcomes using a propensity score matching technique.

In Chapter 5, predictive models to forecast one-year improvement in health-related quality of life (HRQoL) measures were evaluated using the SHAR dataset. Here too, ML models yielded superior AUC values compared to logistic regression and showed "good" to "excellent" discriminative power in classifying significant and substantial clinical improvements, respectively. The top performing model, gradient boosting machine (GBM), identified the three features accounting for over 80% of the ensemble algorithmic function. Primarily, healthier pre-operative baseline PROM and Charnley scores decreased the likelihood of reaching a minimal clinically important difference (MCID),

whilst a sharp increase in this probability was observed for EQ-5D index values of ≤ 0.50, EQ-VAS ≤ 50 and in individuals below 75 years of age. Two previous studies using datasets from Germany (n = 1,843, EQ-5D index AUC = 0.81, EQ-VAS AUC = 0.84) and the United Kingdom (n = 62,429, EQ-VAS index AUC = 0.87), whilst having limitations in terms of clinical applicability and generalisability, reported similar findings (398, 399). The authors in both studies found the GBM model to be the most superior, outperforming logistic regression, but slightly lower AUC values than those reported in Chapter 5. Similarly, both studies identified pre-operative baseline PROM value (EQ-5D index or EQ-VAS) to be the most important feature driving the predictions. Chapter 5, however, acknowledges the need for further research to validate the models and investigate other PROMs such as the Oxford Hip Score (OHS).

The study in Chapter 6 evaluated different machine learning models in predicting the risk of periprosthetic femoral fractures necessitating re-do procedures using the SHAR. Again, ML models showed superior power to traditional statistical approaches with AUCs ranging between 0.79 and 0.86 for revision and re-operation risks between thirty days and one year. Whilst no previous studies attempted to predict periprosthetic femoral fracture risks requiring an intervention, two aforementioned studies forecasted the risks of all-case revision and re-revision following primary THR from a single tertiary referral centre in the United States and reported comparable AUC values (380, 381). However, their limited training dataset size and the absence of disease severity, pre-operative diagnosis, and implant-specific factors in their feature selection may have limited further in-depth explorations. Similar to the previous chapter, the top performing models in Chapter 6 identified the predictors of femoral fractures. These included: cementless fixation type; hip resurfacing arthroplasty; the extremes of femoral head sizes; a pre-operative diagnosis of non-traumatic idiopathic necrosis or inflammatory arthritis; and increasing age, BMI,

and ASA co-morbidity grades. It is noteworthy, however, that the study design did not aim to infer association but to provide insights into these features driving the internal decisions made by the GBM – all of which were thoroughly discussed in the chapter.

Since cementless fixation was identified here and in Chapter 4 to be a predictor of increased revision risk, Chapter 7 examined the effect of cementless fixation on mortality, revision and periprosthetic femoral fracture using the propensity score matching technique and the SHAR dataset. The study reported in a matched cohort of cemented and cementless fixations that uncemented femoral implants in the Swedish population is associated with an improved long-term safety profile but a statistically significant increased risk of revision and periprosthetic femoral fracture. Whilst this is in agreement with previous literature (346, 348, 419-422), the absence of competing variables such as surgical technique and expertise, perioperative care and functional measures warrants the need for prospective trials to investigate the causality.

10.1.2. Part II: Optimising cementless short-stem femoral fixation using biomechanical approaches

Part I identified the augmented risk of periprosthetic femoral fractures (PPFFs) associated with cementless femoral fixation. This comes in common with previous reports from the joint registries in Australia, Sweden and the United Kingdom highlighting that PPFFs are the third most common cause for revision in THR. Whilst not measured in Part I, previous studies alluded to the fact that there remains heterogeneity in surgical technique and experience for cementless stem impaction and implant seating, wherein the impaction energy must be sufficient enough to seat the broach/implant fully but not excessively as to result in fractures (424, 425). As such, the two major studies in Part II of this thesis investigated the effect of femoral impaction energy levels on bone strain behaviour, fracture risk, and implant fixation in cementless short-stem femoral implants.

The study presented in Chapter 8 represents the first study to investigate the effect on impaction energy levels on femoral fracture risk and implant stability using a digital image correlation (DIC) technique and a validated rig that accounts for soft tissue compliance in twenty synthetic bones of low (n = 8) and normal (n = 12) bone mineral density attributes. The laboratory-based experimental study suggested that in low- and normal-density bones and at a fixed impaction angle, a high number of low-energy strikes during femoral broaching and implant seating may provide lower peak strains with similar implant fixation. This phenomenon was thought to be underpinned by the "law of diminishing returns" whereby increased impaction velocity fails to result in an increased implant fixation, beyond which any additional energy produces diminishing additional fixation (444, 455).

To further evaluate the hypothesis generated in this study, Chapter 9 investigated the effect of impact energy on femoral fracture risk, implant stability and fracture threshold in sixteen fresh-frozen cadaveric hips representing the healthy state of bone density. The findings from this study highlighted that there is a statistically significant increased risk of fracture and bone deformation, particularly in the superomedial region of the proximal femur, associated with high-energy strikes during cementless femoral impaction compared to medium and low velocities, whilst implant fixation remains uniform across the varying energy levels. Despite this, several limitations were acknowledged, and are discussed in the following section, and provide a scope for future research in the field prior to translating the findings to the bedside.

10.2. Limitations and future work

This work brought together the fields of health data science, biomechanical engineering, and surgery to understand some of the predictive factors and outcomes after primary

elective THR using machine learning techniques to investigate the effect through which these features could be altered. The depth of knowledge within each domain is vast, and this thesis can only provide limited exploration of these fields and there remains a wide scope for further research.

The first part of this thesis focused on the use of large registry-based datasets, and although they can monitor the safety and failure profiles of different implants, they are not patient safety registries. National joint registries (NJRs) are classified as level 2 (and more recently level 3) on the hierarchical tier system of data (as explored in Section 1.7), lacking thorough collections of radiographs, and in most cases overlooking data related to human and technical factors. Whilst national programmes in the United Kingdom, such as the "Getting It Right First Time", aspire to enhance the quality of care delivered in the NHS, there remains an academic clinical appetite for a holistic and rich national-level database that permits thorough assessments of patient-centred outcomes.

Whilst the concept of NJRs in research can have political implications, the work undertaken in this thesis was conducted at a unit well-known for its work on hip resurfacing arthroplasty (HRA). The principal supervisor is a lead member of a ceramic-on-ceramic HRA (H1) implant, declares grant funding from the Michael Uren Foundation, and has patents and stocks (and stock options) for Embody Orthopaedic Limited, activities outside the confines of this work. The co-supervisor is from an engineering background and has multiple crossovers with industry, is a senior board member of OSSTEC Limited for knee implants and was part of the design team for H1 implant and Additive Instruments Limited. The latter produced a surgical impaction power tool for controlled intra-operative impaction energies, a spin-out company from a doctoral work conducted at Imperial College London and that was recently acquired by Smith & Nephew Limited. Whilst such relationships are difficult to avoid, the author was independently funded by

Imperial College's President's PhD scholarship and is confident that the output produced here refrained from any undue biases or pressures. However, the researcher acknowledges that in the realm of pioneering orthopaedic research, industrial partners form a predominant source of financial support and there remains a need for the research community to establish mechanisms for navigating the funding landscape to mitigate possible external influences.

10.2.1. Part I: Machine learning and outcomes following primary THR

Part I is predicated upon the use of large national observational datasets and machine learning techniques. The size of observations in these studies ensured sufficient sample size and diversity of patient characteristics, surgical practice, and implant factors. This allowed for detailed comparisons of multiple determinants at a country level. In particular, the use of two national registries with comparable findings in Chapters 3 and 4 may point towards a high external validity. Despite this, the limitations of retrospective observational studies were explored in Section 1.7. Whilst each chapter sought to report the relevant existing confounders, there remains a substantial risk of unmeasured confounding.

With regards to confounders related to patient and disease factors, these include: the severity and specificities of co-morbidities; educational and ethnic backgrounds; medication and social histories; previous relevant surgical history; patient expectations; disease severity; PROMs; and functional status. Whilst the latter three were studied in Chapter 5, they were excluded from the other studies as a trade-off for the high level of missingness in their data collection. Nevertheless, it remains crucial to acknowledge that scrutinising registry-derived data and the exclusion of missing variables may inadvertently perpetuate disparities in outcomes. Future studies should seek to investigate the broader dimensions of health inequities pertinent to hip arthroplasty

procedures. In terms of surgical factors, unit and surgeon volume, surgical expertise, and learning curve were not possible to evaluate and are known to form important sources of unmeasured confounding (102, 308). Similarly, data related to peri-operative care, the use of thromboprophylaxis, and more detailed information related to the prosthesis design and brand, other than these included in our study, were also not available. These factors may invariably alter the effect size of the outcomes.

The most significant unmeasured confounders are likely to be surgical expertise, severity of co-morbidities, peri-operative care, and disease severity. Surgical expertise is multifaceted and encompasses technical skills, adaptability in theatres, and decision-making skills, whereas surgical experience is often measured by the volume of procedures performed, grade level or years in practice. In other words, two surgeons may possess an equivalent level of experience but exhibit varying degrees of expertise. Whilst previous studies found a significant effect of surgical experience on patient outcomes in THR (292, 307, 308), the lack of objective metrics to evaluate surgical expertise remains a challenge. Recent advancements in virtual reality appear to accurately measure surgical skill acquisition in THR through assessing procedural sequence, efficiency of movement, and visuospatial accuracy (478). In order to account for such unmeasured confounders, there remains a need to develop and implement standardised, transparent, and objective metrics of procedural skills at a national scale and identify gaps in clinical training that could be improved.

Another source of an unobserved confounding factor is the severity of co-morbidity. National joint registries often use classification systems, such as the American Society of Anaesthesiologists classification or the Charlson Co-morbidity index, to quantify the overall health of patients. Previous literature discussed in Section 1.9 highlighted the association between different medical conditions and mortality and functional outcomes

post-operatively. Likewise, the method, type, and duration of venous thromboembolism (VTE) prophylaxis are established causes to influence mortality outcomes in elective THR (479). The availability and completeness of these variables are unevenly reported in national registries, and future work should seek to address the effect response of these on predictive models. Of these, two-linked RCTs are currently in progress to investigate the strategies for thromboprophylaxis in lower limb immobilisation (480).

In terms of disease severity, the HRQoL predictions presented in Chapter 5 accepted the reduced sample size related to equal variable missingness in HRQoL and disease status and highlighted the feature importance of baseline PROM and Charnley classifications. These variables were excluded from the other studies in this thesis as a trade-off for an increased sample size, which may have limited the predictive power and introduced unobserved bias. On the other hand, the machine learning models in Chapter 5 focused on short term one-year improvements in HRQoL, and it is not clear whether further insights would be identified over the lifetime of THR.

Last but not least, the SHAR surgical practice may be considered heterogenous in terms of fixation methods with over two-thirds of implants being cemented. Therefore, the translation of these findings requires external validation on other national registry datasets such as the NJR in the United Kingdom. Another area that prompts further investigations is the cost-effectiveness of different interventions at a population level, which may yield valuable information to inform policy makers for resource allocation.

10.2.2. Part II: Optimising cementless short-stem femoral fixation using biomechanical approaches

The second part of the thesis concerns the use of a biomechanical lens to understand the bone strain behaviour, fracture risk and implant stability during cementless short-stem

femoral broaching and implant seating. Whilst significant and time-consuming efforts went into setting up a previously validated rig for femoral impaction, systematic specimen preparations, and analyses of the extensive data, the limitations are largely related to the controlled bench-side environment, which were explored in Chapters 8 and 9. These include: strike orientation angle; the absence of strain measurements in the distal femur; the use of single instrument design (Furlong Evolution); and constraints related to the use of digital image correlation and rosette strain gauges. Future studies ought to isolate these specific factors to understand their influence on the studied outcomes.

From a translational perspective, the implementation of such recommendations in surgical practice requires the development of intra-operative instruments that deliver consistent, and reduce variations in, energy levels to reduce the likelihood of fracture. An example of this is represented by Doyle et al.'s work on acetabular cups (444, 447, 448), which developed into a spin-out company, Additive Instruments Limited, translating their bench-side findings to the bed-side through generating a surgical power tool that is set for approval by the Food and Drug Administration (FDA) in the United States (481).

10.3. Conclusions

Non-linear machine learning techniques have a superior binary discriminatory power compared to conventional logistic regression models in predicting post-operative outcomes in THR.

Important features in terms of patient characteristics, surgical practice, and implant design have been identified to improve health-related quality of life measures and reduce mortality, implant failure, and periprosthetic femoral fracture outcomes.

Cementless femoral fixation appears to have a significant effect on improving the long-term safety profile but seem to increase the risks of revision and periprosthetic femoral fractures.

To reduce this risk, laboratory-based experiments seem to suggest that a high number of low-energy strikes during femoral broaching and implant seating may reduce the fracture risk without affecting implant fixation.

Appendices

Appendix 1: Ethics approvals

A1.1. Swedish Hip Arthroplasty Register

Dnr 2019-00957

BESLUT

Sökande forskningshuvudman
Västra Götalandsregionen

Forskare som genomför projektet
Maziar Mohaddes

Projekttitel
Påverkas risken för förnyad operation vid användning av nya implantat eller operationstekniker hos patienter som opereras med en höftprotes?

Aktuell ändring
Ansökan om ändring inkommen 2019-11-17.

Grundansökan godkänd 2017-10-23 av Regionala etikprövningsmyndigheten i Göteborg med diarienummer 804-17.

Etikprövningsmyndigheten beslutar enligt nedan.

BESLUT

Etikprövningsmyndigheten godkänner den forskning som anges i ansökan om ändring.

På Etikprövningsmyndighetens vägnar

Birgitta Henriksson
Ordförande

Beslutet har fattats efter föredragning av vetenskaplig sekreterare Karin Manhem.

Beslutet sänds till
Ansvarig forskare: Maziar Mohaddes

A1.2. Imperial College Healthcare Tissue Bank

Imperial College Healthcare Tissue Bank
Department of Surgery and Cancer
Charing Cross Hospital
Fulham Palace Road
London W6 8RF
Tel: 0203 311 7173
Email: tissuebank@imperial.ac.uk

ICHTB HTA licence: 12275
REC Wales approval: 17/WA/0161

Dear Dr Jonathan Jeffers,

Re: Tissue Bank application number R14088-3A
Project title: Quantifying the importance of soft-tissues for hip joint stability

I am pleased to confirm that the Imperial College Healthcare Tissue Bank has received and approved the following amendments;

Amendment 3A – Adding of the tenth aim of the project and adding work package seven

In order to satisfy HTA tracking requirements, would you please ensure the samples you use are reported to the tissue bank via the online database and sub-collection, along with any publications you produce, that should contain the acknowledgment wording as per the signed MTA.

Yours sincerely,

Professor Iain McNeish
DI, HTA Research Licence.

Imperial College London

Appendix 2: Variable definition list

Variable	Definition	Data type
Age	The age of the patient at the time of primary or re-do procedure	Integer
ASA grade	The class of ASA	Integer; values 1 to 5
BMI	Body mass index at the time of procedure	Decimal, converted to categorical in the studies for clinical relevance
Cement type	Cemented or cementless, and if cemented specify the type of cement used	Integer, around 100 classes exist in the SHAR. Re-classified into five groups: cementless, antibiotic and high viscosity, antibiotic and low viscosity, no antibiotic and high viscosity, no antibiotic and low viscosity
Date	Date of primary or re-do procedure	Datetime
Diagnosis group	Primary pre-operative diagnosis	Integer; primary arthrosis, inflammatory, acute trauma hip fracture, childhood disease, idiopathic necrosis, complication after fracture or trauma, tumour, other secondary arthrosis, acute trauma others, other
Gender	Male or Female	Integer
Hospital type	The type of hospital where procedure was undertaken	Integer: university, county, rural, or private
Incision type	Indicative of surgical approach	Integer, re-classified into five groups: posterior, direct lateral, direct anterior, trochanteric, other
Operation side	Right or left	Integer
Procedure type	Total or hemi hip arthroplasty	Integer
Prosthesis group	All-cemented, all-cementless, hybrid, reverse hybrid or resurfacing	Integer
Cup fixation	Refers to the method or mechanism used to secure or fixate the cup component of the hip prosthesis	Integer, classified into: cemented, cementless or resurfaced
Cup articulation	The material used for the cup articulation surface	Integer, classified into standard metal, resurfaced metal, ceramic, dual mobility (monoblock or modular), standard polyethylene, cross-linked polyethylene, unclear
Cup inner diameter	Refers to the size of the inner opening or bore of the cup component	Integer
Cup modification	Refers to alterations or adjustments made to the cup component	Integer, classified into: standard (non-modified), lipped, dual articular, constrained, unclear
Cup modularity	Refers to the design feature of the cup component, which can be detached from other parts of the implant with a liner (modular), or a single, integrated piece without separable components (monoblock)	Integer, classified into modular or monoblock
Cup resurfacing	Whether the acetabulum was resurfaced with a cup-like component	Integer
Femoral head size	Refers to the diameter of the spherical-shaped top portion of the femoral head that articulates with the acetabular cup	Integer

Femoral fixation	Refers to the method or mechanism used to secure or fixate the stem component of the hip prosthesis	Integer, classified into: cemented, cementless or resurfaced
Femoral modularity	Refers to the design feature of the femoral component, which can be detached from other parts of the implant (modular), or a single, integrated piece without separable components (monoblock)	Integer, classified into modular or monoblock
Femoral resurfacing	Whether the femoral head was resurfaced with a stem-like component	Integer

Appendix 3: Descriptive data from Chapter 3

A3.1 Change in revision rates by year of primary operation in the SHAR

		6-month revision			1-year revision			2-year revision			5-year revision			10-year revision		
Year	N total	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§
2000-2008	11,155	91	0.82	1.0	133	1.19	1.0	213	1.91	1.0	338	3.03	1.0	627	5.62	1.0
2009	13,027	118	0.91	1.1 (0.8-1.4)	165	1.27	1.0 (0.8-1.3)	240	1.84	0.9 (0.8-1.1)	378	2.90	0.9 (0.8-1.1)	642	4.93	0.8 (0.7-0.9)*
2010	13,546	113	0.83	1.0 (0.7-1.3)	156	1.15	0.9 (0.7-1.2)	225	1.66	0.8 (0.7-1.0)	394	2.91	0.9 (0.8-1.1)	593	4.38	0.7 (0.6-0.8)***
2011	13,697	146	1.07	1.3 (1.0-1.6)*	190	1.39	1.1 (0.9-1.4)	267	1.95	1.0 (0.8-1.2)	420	3.07	1.0 (0.8-1.1)	537	3.92	0.6 (0.6-0.7)***
2012	13,946	164	1.18	1.4 (1.1-1.8)**	201	1.44	1.2 (0.9-1.4)	285	2.04	1.0 (0.8-1.2)	454	3.26	1.0 (0.9-1.2)	520	3.73	0.6 (0.5-0.7)***
2013	13,989	166	1.19	1.4 (1.1-1.8)**	210	1.50	1.2 (1.0-1.5)*	273	1.95	1.0 (0.8-1.2)	435	3.11	1.0 (0.8-1.1)	454	3.25	0.5 (0.5-0.6)***
2014	14,226	182	1.28	1.5 (1.2-1.9)***	229	1.61	1.3 (1.0-1.6)**	301	2.12	1.0 (0.9-1.3)	435	3.06	1.0 (0.8-1.1)	435	3.06	0.5 (0.4-0.6)***
2015	14,403	165	1.15	1.3 (1.0-1.8)*	212	1.47	1.2 (0.9-1.5)	279	1.94	1.0 (0.8-1.2)	367	2.55	0.8 (0.7-0.9)*	367	2.55	0.4 (0.3-0.5)***
2016	14,995	217	1.45	1.7 (1.3-2.2)***	280	1.87	1.5 (1.2-1.9)***	376	2.51	1.3 (1.1-1.5)**	406	2.71	0.8 (0.7-1.0)	406	2.71	0.4 (0.4-0.5)***
2017	15,512	209	1.35	1.6 (1.2-2.0)***	253	1.63	1.3 (1.1-1.6)**	294	1.90	0.9 (0.8-1.1)	294	1.90	0.6 (0.5-0.7)***	294	1.90	0.3 (0.2-0.3)***
2018	16,099	244	1.52	1.83 (1.4-2.3)***	258	1.60	1.3 (1.0-1.6)**	258	1.60	0.8 (0.6-1.0)	258	1.60	0.5 (0.4-0.6)***	258	1.60	0.2 (0.2-0.3)***

§ Adjusted for Age and Gender; * p<0.05, ** p<0.01, *** p<0.001
KM = Kaplan-Meier; HR = Hazard Ratio; CI = Confidence Interval

A3.2 Change in re-operation rates by year of primary operation in the SHAR

		6-month re-operation			1-year re-operation			2-year re-operation			5-year re-operation			10-year re-operation		
Year	N total	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§	N	KM (%)	HR (95% CI)§
2000-2008	11,155	196	1.76	1.0	251	2.25	1.0	342	3.07	1.0	502	4.50	1.0	857	7.68	1.0
2009	13,027	225	1.73	0.9 (0.8-1.1)	282	2.16	0.9 (0.8-1.1)	379	2.91	0.9 (0.8-1.0)	546	4.19	0.9 (0.8-1.0)	867	6.66	0.8 (0.7-0.9)**
2010	13,546	205	1.51	0.8 (0.7-1.0)	273	2.02	0.8 (0.7-1.0)	380	2.81	0.9 (0.7-1.0)	596	4.40	0.9 (0.8-1.0)	830	6.13	0.7 (0.7-0.8)***
2011	13,697	247	1.80	1.0 (0.8-1.2)	311	2.27	1.0 (0.8-1.1)	414	3.02	0.9 (0.8-1.1)	625	4.56	1.0 (0.8-1.1)	760	5.55	0.7 (0.6-0.7)***
2012	13,946	275	1.97	1.1 (0.9-1.3)	344	2.47	1.0 (0.9-1.2)	471	3.38	1.0 (0.9-1.2)	691	4.95	1.1 (0.9-1.2)	768	5.51	0.7 (0.6-0.7)***
2013	13,989	267	1.91	1.0 (0.8-1.2)	323	2.31	1.0 (0.8-1.2)	414	2.96	0.9 (0.8-1.1)	601	4.30	0.9 (0.8-1.0)	626	4.47	0.5 (0.5-0.6)***
2014	14,226	263	1.85	1.0 (0.8-1.2)	328	2.31	1.0 (0.8-1.1)	430	3.02	0.9 (0.8-1.1)	594	4.18	0.9 (0.8-1.0)	594	4.18	0.5 (0.4-0.5)***
2015	14,403	260	1.81	1.0 (0.8-1.2)	327	2.27	0.9 (0.8-1.1)	413	2.87	0.9 (0.8-1.0)	521	3.62	0.7 (0.7-0.9)***	521	3.62	0.4 (0.4-0.5)***
2016	14,995	307	2.05	1.1 (0.9-1.3)	384	2.56	1.1 (0.9-1.3)	489	3.26	1.0 (0.9-1.2)	526	3.51	0.7 (0.6-0.8)***	526	3.51	0.4 (0.4-0.5)***
2017	15,512	276	1.78	0.9 (0.8-1.1)	336	2.17	0.9 (0.8-1.1)	388	2.50	0.8 (0.6-0.9)**	388	2.50	0.5 (0.4-0.6)***	388	2.50	0.3 (0.2-0.3)***
2018	16,099	290	1.80	1.0 (0.8-1.2)	308	1.91	0.8 (0.7-0.9)*	308	1.91	0.6 (0.5-0.7)***	308	1.91	0.4 (0.3-0.4)***	308	1.91	0.2 (0.2-0.2)***

§ Adjusted for Age and Gender; * p<0.05, ** p<0.01, *** p<0.001
KM = Kaplan-Meier; HR = Hazard Ratio; CI = Confidence Interval

Bibliography

1. McMinn D, Snell KIE, Daniel J, Treacy R, Pynsent P, Riley R. Mortality and implant revision rates of hip arthroplasty in patients with osteoarthritis: registry based cohort study. Bmj. 2012;344.
2. Hughes RE, Batra A, Hallstrom BR. Arthroplasty registries around the world: valuable sources of hip implant revision risk data. Curr Rev Musculoskelet Med. 2017;10(2):240-52.
3. Okafor CE, Nghiem S, Byrnes J. Are joint replacement registries associated with burden of revision changes? A real-world panel data regression analysis. BMJ Open. 2023;13(1):e063472.
4. Baker PN, Jeyapalan R, Jameson SS. The value of national arthroplasty registry data in 2023. Bone Joint J. 2023;105(4):356-60.
5. The Swedish Arthroplasty Register. Annual Report.; 2022.
6. National Joint Registry - Annual Report.; 2022.
7. Australian Orthopaedic Association National Joint Replacement Registry - Annual Report. 2022.
8. Whitehouse SL, Bolland BJRF, Howell JR, Crawford RW, Timperley AJ. Mortality Following Hip Arthroplasty—Inappropriate Use of National Joint Registry (NJR) Data. The Journal of Arthroplasty. 2014;29(9):1827-34.
9. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. The New England journal of medicine. 2016;375(13):1216.
10. Farooq H, Deckard ER, Ziemba-Davis M, Madsen A, Meneghini RM. Predictors of Patient Satisfaction Following Primary Total Knee Arthroplasty: Results from a Traditional Statistical Model and a Machine Learning Algorithm. The Journal of Arthroplasty. 2020;35(11):3123-30.
11. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. The New England journal of medicine. 2017;376(26):2507.
12. Hunt LP, Ben-Shlomo Y, Clark EM, Dieppe P, Judge A, MacGregor AJ, et al. 90-day mortality after 409 096 total hip replacements for osteoarthritis, from the National Joint Registry for England and Wales: a retrospective analysis. The Lancet. 2013;382(9898):1097-104.
13. Kendal AR, Prieto-Alhambra D, Arden NK, Carr A, Judge A. Mortality rates at 10 years after metal-on-metal hip resurfacing compared with total hip replacement in England: retrospective cohort analysis of hospital episode statistics. Bmj. 2013;347.
14. Hunt LP, Whitehouse MR, Howard PW, Ben-Shlomo Y, Blom AW. Using long term mortality to determine which perioperative risk factors of mortality following hip and knee replacement may be causal. Scientific Reports. 2018;8(1):15026.
15. Hunter DJ, March L, Chew M. Osteoarthritis in 2020 and beyond: a Lancet Commission. The Lancet. 2020;396(10264):1711-2.
16. Hamood R, Tirosh M, Fallach N, Chodick G, Eisenberg E, Lubovsky O. Prevalence and incidence of osteoarthritis: A population-based retrospective cohort study. Journal of clinical medicine. 2021;10(18):4282.
17. Litwic A, Edwards MH, Dennison EM, Cooper C. Epidemiology and burden of osteoarthritis. British medical bulletin. 2013;105(1):185-99.

18. Safiri S, Kolahi A-A, Smith E, Hill C, Bettampadi D, Mansournia MA, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. Annals of the rheumatic diseases. 2020;79(6):819-28.
19. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of hip symptoms and radiographic and symptomatic hip osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. The Journal of rheumatology. 2009;36(4):809-15.
20. Nelson AE, Hu D, Arbeeva L, Alvarez C, Cleveland RJ, Schwartz TA, et al. Point prevalence of hip symptoms, radiographic, and symptomatic OA at five time points: The Johnston County Osteoarthritis Project, 1991–2018. Osteoarthritis and cartilage open. 2022;4(2):100251.
21. Register B, Pennock AT, Ho CP, Strickland CD, Lawand A, Philippon MJ. Prevalence of Abnormal Hip Findings in Asymptomatic Participants:A Prospective, Blinded Study. The American Journal of Sports Medicine. 2012;40(12):2720-4.
22. Guevara CJ, Pietrobon R, Carothers JT, Olson SA, Vail TP. Comprehensive morphologic evaluation of the hip in patients with symptomatic labral tear. Clinical Orthopaedics and Related Research®. 2006;453:277-85.
23. Heerey JJ, Srinivasan R, Agricola R, Smith A, Kemp JL, Pizzari T, et al. Prevalence of early hip OA features on MRI in high-impact athletes. The femoroacetabular impingement and hip osteoarthritis cohort (FORCe) study. Osteoarthritis and Cartilage. 2021;29(3):323-34.
24. Fujii Y, Liu L, Yagasaki L, Inotsume M, Chiba T, Asahara H. Cartilage Homeostasis and Osteoarthritis. International Journal of Molecular Sciences. 2022;23(11):6316.
25. Cooper C, Inskip H, Croft P, Campbell L, Smith G, Mclearn M, et al. Individual Risk factors for Hip Osteoarthritis: Obesity, Hip Injury and Physical Activity. American Journal of Epidemiology. 1998;147(6):516-22.
26. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med. 2010;26(3):355-69.
27. Lawrence JS, Bremner JM, Bier F. Osteo-arthrosis. Prevalence in the population and relationship between symptoms and x-ray changes. Ann Rheum Dis. 1966;25(1):1-24.
28. Mourad C, Vande Berg B. Osteoarthritis of the hip: is radiography still needed? Skeletal Radiology. 2022.
29. van Berkel AC, Schiphof D, Waarsing JH, Runhaar J, van Ochten JM, Bindels PJE, et al. Course of pain and fluctuations in pain related to suspected early hip osteoarthritis: the CHECK study. Fam Pract. 2022;39(6):1041-8.
30. Matharu G, Culliford D, Blom A, Judge A. Projections for primary hip and knee replacement surgery up to the year 2060: an analysis based on data from The National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. The Annals of The Royal College of Surgeons of England. 2022;104(6):443-8.
31. Singh JA, Yu S, Chen L, Cleveland JD. Rates of total joint replacement in the United States: future projections to 2020–2040 using the national inpatient sample. The Journal of rheumatology. 2019;46(9):1134-40.
32. Dakin H, Eibich P, Beard D, Gray A, Price A, team AS. The use of patient-reported outcome measures to guide referral for hip and knee replacement: Part 2–a cost-effectiveness analysis. The Bone and Joint Journal. 2020;102.
33. Price AJ, Kang S, Cook JA, Dakin H, Blom A, Arden N, et al. The use of patient-reported outcome measures to guide referral for hip and knee arthroplasty: part 1: the development of an evidence-based model linking preoperative score to the probability of gaining benefit from surgery. The Bone & Joint Journal. 2020;102(7):941-9.
34. NICE. Osteoarthritis in over 16s: diagnosis and management. NICE guideline.; 2022.

35. Dakin HA, Eibich P, Gray A, Smith J, Barker KL, Beard D, et al. Who gets referred for knee or hip replacement? A theoretical model of the potential impact of evidence-based referral thresholds using data from a retrospective review of clinic records from an English musculoskeletal referral hub. BMJ Open. 2020;10(7):e028915.
36. 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;61(589):e459-68.
37. Postler AE, Lützner C, Goronzy J, Lange T, Deckert S, Günther KP, et al. When are patients with osteoarthritis referred for surgery? Best Practice & Research Clinical Rheumatology. 2023:101835.
38. Bannuru RR, Osani M, Vaysbrot E, Arden N, Bennell K, Bierma-Zeinstra S, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis and cartilage. 2019;27(11):1578-89.
39. Holden MA, Metcalf B, Lawford BJ, Hinman RS, Boyd M, Button K, et al. Recommendations for the delivery of therapeutic exercise for people with knee and/or hip osteoarthritis. An international consensus study from the OARSI Rehabilitation Discussion Group. Osteoarthritis and Cartilage. 2023;31(3):386-96.
40. Meng Z, Liu J, Zhou N. Efficacy and safety of the combination of glucosamine and chondroitin for knee osteoarthritis: a systematic review and meta-analysis. Arch Orthop Trauma Surg. 2023;143(1):409-21.
41. Roman-Blas JA, Castañeda S, Sánchez-Pernaute O, Largo R, Herrero-Beaumont G. Combined Treatment With Chondroitin Sulfate and Glucosamine Sulfate Shows No Superiority Over Placebo for Reduction of Joint Pain and Functional Impairment in Patients With Knee Osteoarthritis: A Six-Month Multicenter, Randomized, Double-Blind, Placebo-Controlled Clinical Trial. Arthritis Rheumatol. 2017;69(1):77-85.
42. Towheed T, Maxwell L, Anastassiades TP, Shea B, Houpt J, Welch V, et al. Glucosamine therapy for treating osteoarthritis. Cochrane database of systematic reviews. 2005(2).
43. Surgeons AAoO. Management of Osteoarthritis of the Hip Evidence-Based Clinical Practice Guideline. 2017.
44. Lai WC, Arshi A, Wang D, Seeger LL, Motamedi K, Levine BD, et al. Efficacy of intraarticular corticosteroid hip injections for osteoarthritis and subsequent surgery. Skeletal Radiol. 2018;47(12):1635-40.
45. McCabe PS, Maricar N, Parkes MJ, Felson DT, O'Neill TW. The efficacy of intra-articular steroids in hip osteoarthritis: a systematic review. Osteoarthritis Cartilage. 2016;24(9):1509-17.
46. Gazendam A, Ekhtiari S, Bozzo A, Phillips M, Bhandari M. Intra-articular saline injection is as effective as corticosteroids, platelet-rich plasma and hyaluronic acid for hip osteoarthritis pain: a systematic review and network meta-analysis of randomised controlled trials. Br J Sports Med. 2021;55(5):256-61.
47. Eymard F, Maillet B, Lellouche H, Mellac-Ducamp S, Brocq O, Loeuille D, et al. Predictors of response to viscosupplementation in patients with hip osteoarthritis: results of a prospective, observational, multicentre, open-label, pilot study. BMC Musculoskeletal Disorders. 2017;18(1):1-8.
48. Quinn RH, Murray J, Pezold R, Hall Q. Management of Osteoarthritis of the Hip. J Am Acad Orthop Surg. 2018;26(20):e434-e6.
49. Basedow M, Williams H, Shanahan EM, Runciman WB, Esterman A. Australian GP management of osteoarthritis following the release of the RACGP guideline for the non-surgical management of hip and knee osteoarthritis. BMC Res Notes. 2015;8:536.

50. Dong Y, Zhang B, Yang Q, Zhu J, Sun X. The effects of platelet-rich plasma injection in knee and hip osteoarthritis: a meta-analysis of randomized controlled trials. Clin Rheumatol. 2021;40(1):263-77.
51. Mardones R, Jofré CM, Tobar L, Minguell JJ. Mesenchymal stem cell therapy in the treatment of hip osteoarthritis. J Hip Preserv Surg. 2017;4(2):159-63.
52. McIntyre JA, Jones IA, Han B, Vangsness CT, Jr. Intra-articular Mesenchymal Stem Cell Therapy for the Human Joint: A Systematic Review. Am J Sports Med. 2018;46(14):3550-63.
53. Rampal S, Jaiman A, Tokgöz MA, Arumugam G, Sivananthan S, Singh RSJ, et al. A review of the efficacy of intraarticular hip injection for patients with hip osteoarthritis: To inject or not to inject in hip osteoarthritis? Jt Dis Relat Surg. 2022;33(1):255-62.
54. Rodriguez-Fontan F, Piuzzi NS, Kraeutler MJ, Pascual-Garrido C. Early Clinical Outcomes of Intra-Articular Injections of Bone Marrow Aspirate Concentrate for the Treatment of Early Osteoarthritis of the Hip and Knee: A Cohort Study. Pm r. 2018;10(12):1353-9.
55. Park KD, Kim TK, Bae BW, Ahn J, Lee WY, Park Y. Ultrasound guided intra-articular ketorolac versus corticosteroid injection in osteoarthritis of the hip: a retrospective comparative study. Skeletal Radiol. 2015;44(9):1333-40.
56. Jurgensmeier K, Jurgensmeier D, Kunz DE, Fuerst PG, Warth LC, Daines SB. Intra-articular Injections of the Hip and Knee With Triamcinolone vs Ketorolac: A Randomized Controlled Trial. J Arthroplasty. 2021;36(2):416-22.
57. Waluyo Y, Artika SR, Wahyuni IN, Gunawan AMAK, Zainal ATF. Efficacy of Prolotherapy for Osteoarthritis: A Systematic Review. Journal of Rehabilitation Medicine. 2023;55.
58. Andronic O, Claydon-Mueller LS, Cubberley R, Karczewski D, Sunil-Kumar KH, Khanduja V. Inconclusive and contradictory evidence for outcomes after hip arthroscopy in patients with Femoroacetabular impingement and osteoarthritis of Tönnis grade 2 or greater: a systematic review. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2022;38(7):2307-18. e1.
59. Cross GWV, Sobti AS, Khan T. Hip arthroscopy in osteoarthritis: Is it an option? Journal of Clinical Orthopaedics and Trauma. 2021;22:101617.
60. Schwabe MT, Clohisy JC, Cheng AL, Pascual-Garrido C, Harris-Hayes M, Hunt DM, et al. Short-term Clinical Outcomes of Hip Arthroscopy Versus Physical Therapy in Patients With Femoroacetabular Impingement: A Systematic Review and Meta-analysis of Randomized Controlled Trials. Orthop J Sports Med. 2020;8(11):2325967120968490.
61. Martin SD, Abraham PF, Varady NH, Nazal MR, Conaway W, Quinlan NJ, et al. Hip Arthroscopy Versus Physical Therapy for the Treatment of Symptomatic Acetabular Labral Tears in Patients Older Than 40 Years: A Randomized Controlled Trial. Am J Sports Med. 2021;49(5):1199-208.
62. Piuzzi NS, Slullitel PAI, Bertona A, Oñativia IJ, Albergo I, Zanotti G, et al. Hip Arthroscopy in Osteoarthritis: A Systematic Review of the Literature. HIP International. 2016;26(1):8-14.
63. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. The Lancet. 2007;370(9597):1508-19.
64. Smith-Petersen M. Evolution of mould arthroplasty of the hip joint. The Journal of Bone & Joint Surgery British Volume. 1948;30(1):59-75.
65. Wiles P. The surgery of the osteo-arthritic hip. British Journal of Surgery. 1958;45(193):488-97.
66. Bota NC, Nistor DV, Caterev S, Todor A. Historical overview of hip arthroplasty: From humble beginnings to a high-tech future. Orthop Rev (Pavia). 2021;13(1):8773.

67. McKellop H, Park S-H, Chiesa R, Doorn P, Lu B, Normand P, et al. In Vivo Wear of 3 Types of Metal on Metal Hip Prostheses During 2 Decades of Use. Clinical Orthopaedics and Related Research (1976-2007). 1996;329:S128-S40.
68. Charnley J. Arthroplasty of the hip. A new operation. Lancet. 1961;1(7187):1129-32.
69. Charnley J. Anchorage of the femoral head prosthesis to the shaft of the femur. The Journal of bone and joint surgery British volume. 1960;42(1):28-30.
70. Caton J, Prudhon JL. Over 25 years survival after Charnley's total hip arthroplasty. Int Orthop. 2011;35(2):185-8.
71. Harris WH, Schiller AL, Scholler J, Freiberg RA, Scott R. Extensive localized bone resorption in the femur following total hip replacement. JBJS. 1976;58(5):612-8.
72. Gerard Y. Hip arthroplasty by matching cups. Clinical Orthopaedics and Related Research (1976-2007). 1978;134:25-35.
73. Freeman M, Cameron H, Brown G. Cemented double cup arthroplasty of the hip: a 5 year experience with the ICLH prosthesis. Clinical orthopaedics and related research. 1978(134):45-52.
74. McMinn D, Treacy R, Lin K, Pynsent P. Metal on metal surface replacement of the hip: experience of the McMinn prosthesis. Clinical Orthopaedics and Related Research (1976-2007). 1996;329:S89-S98.
75. Cobb JP, Clarke SJ, Jeffers J, Wozencroft R, Halewood C, Amis AA. ANATOMIC CERAMIC HIP RESURFACING ARTHROPLASTY: FIRST IN HUMAN TRIALS. Orthopaedic Proceedings. 2018;100-B(SUPP_1):29-.
76. Bucholz RW. Indications, techniques and results of total hip replacement in the united states. Revista Médica Clínica Las Condes. 2014;25(5):756-9.
77. Lombardi B, Paci M, Nannetti L, Moretti S, Maritato M, Benelli G. Total hip arthroplasty after hip fracture or osteoarthritis: are there differences in characteristics and outcomes in the early rehabilitative stage? Orthopaedic Nursing. 2014;33(1):43-7.
78. Stirton JB, Maier JC, Nandi S. Total hip arthroplasty for the management of hip fracture: A review of the literature. J Orthop. 2019;16(2):141-4.
79. Patel N, Golwala P. Approaches for Total Hip Arthroplasty: A Systematic Review. Cureus. 2023;15(2).
80. Onyemaechi N, Anyanwu E, Obikili E, Ekezie J. Anatomical basis for surgical approaches to the hip. Ann Med Health Sci Res. 2014;4(4):487-94.
81. Hiran A. Surgical Approaches to the Hip Joint and Its Clinical Implications in Adult Hip Arthroplasty. In: Plamen K, editor. Arthroplasty. Rijeka: IntechOpen; 2013. p. Ch. 1.
82. Petis S, Howard JL, Lanting BL, Vasarhelyi EM. Surgical approach in primary total hip arthroplasty: anatomy, technique and clinical outcomes. Canadian Journal of Surgery. 2015;58(2):128.
83. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. BMJ Quality & Safety. 1996;5(2):81-8.
84. Docter S, Philpott HT, Godkin L, Bryant D, Somerville L, Jennings M, et al. Comparison of intra and post-operative complication rates among surgical approaches in Total Hip Arthroplasty: A systematic review and meta-analysis. Journal of Orthopaedics. 2020;20:310-25.
85. Miller LE, Gondusky JS, Kamath AF, Boettner F, Wright J, Bhattacharyya S. Influence of surgical approach on complication risk in primary total hip arthroplasty: Systematic review and meta-analysis. Acta orthopaedica. 2018;89(3):289-94.
86. Su EP. Retraction: Post-operative neuropathy after total hip arthroplasty. The Bone & Joint Journal. 2017;99-B(1_Supple_A):46-9.
87. McFarland B, Osborne G. Approach to the hip: a suggested improvement on Kocher's method. The Journal of Bone and Joint Surgery British volume. 1954;36(3):364-7.

88. Hardinge K. The direct lateral approach to the hip. The Journal of bone and joint surgery British volume. 1982;64(1):17-9.
89. Smith-Petersen M. Approach to and exposure of the hip joint for mold arthroplasty. JBJS. 1949;31(1):40-6.
90. Oinuma K, Eingartner C, Saito Y, Shiratsuchi H. Total hip arthroplasty by a minimally invasive, direct anterior approach. Operative Orthopädie und Traumatologie. 2007;19:310-26.
91. Post ZD, Orozco F, Diaz-Ledezma C, Hozack WJ, Ong A. Direct anterior approach for total hip arthroplasty: indications, technique, and results. JAAOS-Journal of the American Academy of Orthopaedic Surgeons. 2014;22(9):595-603.
92. Cichos KH, McGwin Jr G, Boyd B, Patel SS, Cao AQ, Jordan EM, et al. Direct Anterior Approach Total Hip Arthroplasty Is Associated With Reduced 1-Year Mortality and Surgical Complications After Femoral Neck Fracture. The Journal of Arthroplasty. 2023.
93. Cichos KH, McGwin G, Boyd B, Cichos KH, Patel SS, Cao AQ, et al. Direct Anterior Approach Total Hip Arthroplasty Is Associated With Reduced 1-Year Mortality and Surgical Complications After Femoral Neck Fracture. The Journal of Arthroplasty. 2023.
94. Ang JJM, Onggo JR, Stokes CM, Ambikaipalan A. Comparing direct anterior approach versus posterior approach or lateral approach in total hip arthroplasty: a systematic review and meta-analysis. European Journal of Orthopaedic Surgery & Traumatology. 2023.
95. Awad ME, Farley BJ, Mostafa G, Saleh KJ. Direct anterior approach has short-term functional benefit and higher resource requirements compared with the posterior approach in primary total hip arthroplasty: a meta-analysis of functional outcomes and cost. The bone & joint journal. 2021;103(6):1078-87.
96. Angerame MR, Fehring TK, Masonis JL, Mason JB, Odum SM, Springer BD. Early failure of primary total hip arthroplasty: is surgical approach a risk factor? The Journal of arthroplasty. 2018;33(6):1780-5.
97. Taunton MJ, Mason JB, Odum SM, Springer BD. Direct anterior total hip arthroplasty yields more rapid voluntary cessation of all walking aids: a prospective, randomized clinical trial. The Journal of arthroplasty. 2014;29(9):169-72.
98. Taunton MJ, Trousdale RT, Sierra RJ, Kaufman K, Pagnano MW. John charnley award: randomized clinical trial of direct anterior and miniposterior approach THA: which provides better functional recovery? Clinical orthopaedics and related research. 2018;476(2):216.
99. Sibia US, Turner TR, MacDonald JH, King PJ. The impact of surgical technique on patient reported outcome measures and early complications after total hip arthroplasty. The Journal of arthroplasty. 2017;32(4):1171-5.
100. Rodriguez JA, Deshmukh AJ, Rathod PA, Greiz ML, Deshmane PP, Hepinstall MS, et al. Does the direct anterior approach in THA offer faster rehabilitation and comparable safety to the posterior approach? Clinical Orthopaedics and Related Research®. 2014;472:455-63.
101. Peters RM, van Beers LW, van Steenbergen LN, Wolkenfelt J, Ettema HB, Ten Have BL, et al. Similar superior patient-reported outcome measures for anterior and posterolateral approaches after total hip arthroplasty: postoperative patient-reported outcome measure improvement after 3 months in 12,774 primary total hip arthroplasties using the anterior, anterolateral, straight lateral, or posterolateral approach. The Journal of arthroplasty. 2018;33(6):1786-93.
102. Nairn L, Gyemi L, Gouveia K, Ekhtiari S, Khanna V. The learning curve for the direct anterior total hip arthroplasty: a systematic review. International Orthopaedics. 2021;45:1971-82.

103. Peters RM, ten Have BLEF, Rykov K, van Steenbergen L, Putter H, Rutgers M, et al. The learning curve of the direct anterior approach is 100 cases: an analysis based on 15,875 total hip arthroplasties in the Dutch Arthroplasty Register. Acta Orthopaedica. 2022;93:775-82.
104. Watson-Jones R. Fractures of the neck of the femur. British Journal of Surgery. 1936;23(92):787-808.
105. Mueller M, editor Total hip replacement without trochanteric osteotomy. The Hip. Proc, Second Open Scientific Meeting of the Hip Society, St Louis, The CV Mosby Co; 1974.
106. Charnley J. Low friction arthroplasty of the hip: theory and practice. Berlin: Springer-Verlag; 1979. p. 376.
107. Breusch S, Malchau H, Breusch SJ. Operative Steps: Femur. The Well-Cemented Total Hip Arthroplasty: Theory and Practice. 2005:125-40.
108. Srinivasan S, Shah R, Rayan F, Ensor D, Sambhwani S, Menon DK. Centralizing the cemented exeter femoral stem using the direct lateral approach: surgical tips and radiological evaluation. Arthroplasty Today. 2020;6(4):755-60.
109. DeLee J, Charnley J. Radiological demarcation of cemented sockets in hip arthroplasty. Clin Orthop Relat Res. 1976;121:20-32.
110. Gruen TA, McNeice GM, Amstutz HC. "Modes of failure" of cemented stem-type femoral components: a radiographic analysis of loosening. Clin Orthop Relat Res. 1979(141):17-27.
111. Duncan CP, Masri BA. Fractures of the femur after hip replacement. Instr Course Lect. 1995;44:293-304.
112. Davenport D, Hutt JR, Mitchell PA, Trompeter A, Kendoff D, Sandiford NA. Management of peri-prosthetic fractures around total hip arthroplasty: a contemporary review of surgical options. Annals of Joint. 2018;3:65-.
113. Brady OH, Garbuz DS, Masri BA, Duncan CP. The reliability and validity of the Vancouver classification of femoral fractures after hip replacement. J Arthroplasty. 2000;15(1):59-62.
114. Wagner M, Wagner H. Preliminary results of uncemented metal on metal stemmed and resurfacing hip replacement arthroplasty. Clinical Orthopaedics and Related Research (1976-2007). 1996;329:S78-S88.
115. McMinn DJ. Modern hip resurfacing: Springer Science & Business Media; 2009.
116. Al-Jabri T, Ridha M, McCulloch RA, Kayani B, Arif A, Habad M, et al. Hip Resurfacing Arthroplasty: Past, Present and Future. Orthop Rev (Pavia). 2023;15:77745.
117. Coulter G, Young DA, Dalziel RE, Shimmin AJ. Birmingham hip resurfacing at a mean of ten years. The Journal of Bone & Joint Surgery British Volume. 2012;94-B(3):315-21.
118. Treacy RBC, McBryde CW, Shears E, Pynsent PB. Birmingham hip resurfacing. The Journal of Bone & Joint Surgery British Volume. 2011;93-B(1):27-33.
119. Holland JP, Langton DJ, Hashmi M. Ten-year clinical, radiological and metal ion analysis of the Birmingham Hip Resurfacing. The Journal of Bone & Joint Surgery British Volume. 2012;94-B(4):471-6.
120. Murray DW, Grammatopoulos G, Pandit H, Gundle R, Gill HS, McLardy-Smith P. The ten-year survival of the Birmingham hip resurfacing. The Journal of Bone & Joint Surgery British Volume. 2012;94-B(9):1180-6.
121. Amstutz HC, Le Duff MJ, Campbell PA, Gruen TA, Wisk LE. Clinical and radiographic results of metal-on-metal hip resurfacing with a minimum ten-year follow-up. JBJS. 2010;92(16):2663-71.
122. Haddad FS, Thakrar RR, Hart AJ, Skinner JA, Nargol AVF, Nolan JF, et al. Metal-on-metal bearings. The Journal of Bone & Joint Surgery British Volume. 2011;93-B(5):572-9.

123. Maniatopoulos G, Hopkins C, Joyce TJ, Brittain K. Framing the failure of medical implants: Media representations of the ASR hip replacements in the UK. Health Expectations. 2019;22(3):518-27.
124. Clough EJ, Clough TM. Metal on metal hip resurfacing arthroplasty: Where are we now? J Orthop. 2021;23:123-7.
125. Pandit H, Glyn-Jones S, McLardy-Smith P, Gundle R, Whitwell D, Gibbons C, et al. Pseudotumours associated with metal-on-metal hip resurfacings. The Journal of Bone & Joint Surgery British Volume. 2008;90(7):847-51.
126. Ollivere B, Darrah C, Barker T, Nolan J, Porteous M. Early clinical failure of the Birmingham metal-on-metal hip resurfacing is associated with metallosis and soft-tissue necrosis. The Journal of Bone & Joint Surgery British Volume. 2009;91(8):1025-30.
127. Kwon Y-M, Glyn-Jones S, Simpson D, Kamali A, McLardy-Smith P, Gill H, et al. Analysis of wear of retrieved metal-on-metal hip resurfacing implants revised due to pseudotumours. The Journal of Bone & Joint Surgery British Volume. 2010;92(3):356-61.
128. Langton D, Joyce T, Jameson S, Lord J, Van Orsouw M, Holland J, et al. Adverse reaction to metal debris following hip resurfacing: the influence of component type, orientation and volumetric wear. The Journal of Bone & Joint Surgery British Volume. 2011;93(2):164-71.
129. Girard J, Lons A, Ramdane N, Putman S. Hip resurfacing before 50years of age: A prospective study of 979 hips with a mean follow-up of 5.1years. Orthopaedics & Traumatology: Surgery & Research. 2018;104(3):295-9.
130. Matharu GS, McBryde CW, Pynsent WB, Pynsent PB, Treacy RBC. The outcome of the Birmingham Hip Resurfacing in patients aged < 50 years up to 14 years post-operatively. The Bone & Joint Journal. 2013;95-B(9):1172-7.
131. Ford MC, Hellman MD, Kazarian GS, Clohisy JC, Nunley RM, Barrack RL. Five to Ten-Year Results of the Birmingham Hip Resurfacing Implant in the U.S.: A Single Institution's Experience. J Bone Joint Surg Am. 2018;100(21):1879-87.
132. McBryde CW, Prakash R, Haddad FS. Hip resurfacing. The Bone & Joint Journal. 2023;105-B(5):467-70.
133. Logishetty K, van Arkel RJ, Ng KCG, Muirhead-Allwood SK, Cobb JP, Jeffers JRT. Hip capsule biomechanics after arthroplasty: the effect of implant, approach, and surgical repair. Bone Joint J. 2019;101-b(4):426-34.
134. Maslivec A, Ng KCG, Cobb J. BILATERAL HIP RESURFACING ARTHROPLASTY RESTORES GAIT FUNCTION AT INCREASING WALKING SPEEDS. Orthopaedic Proceedings. 2021;103-B(SUPP_16):54-.
135. Maillot C, Auvinet E, Harman C, Cobb J, Rivière C. Hip resurfacing generates a more physiological gait than total hip replacement: A case-control study. Orthopaedics & Traumatology: Surgery & Research. 2020;106(3):527-34.
136. Van Der Straeten C, Gross TP, Amstutz H, Brooks PJ, Samuel LT, Su EP, et al. Hip resurfacing arthroplasty in young patients: international high-volume centres' report on the outcome of 11,382 metal-on-metal hip resurfacing arthroplasties in patients ⩽50 years at surgery. HIP International. 2022;32(3):353-62.
137. Amstutz HC, Le Duff MJ. Sex-specific risk factors determine the survivorship of female and male patients after metal-on-metal hip resurfacing. Hip Int. 2020;30(3):309-18.
138. McMinn DJW, Daniel J, Ziaee H, Pradhan C. Indications and results of hip resurfacing. International Orthopaedics. 2011;35(2):231-7.
139. Nunley RM, Della Valle CJ, Barrack RL. Is patient selection important for hip resurfacing? Clinical orthopaedics and related research. 2009;467:56-65.
140. Anglin C, Masri BA, Tonetti J, Hodgson AJ, Greidanus NV. Hip resurfacing femoral neck fracture influenced by valgus placement. Clinical Orthopaedics and Related Research (1976-2007). 2007;465:71-9.

141. Slullitel P, Grammatopoulos G, Calistr A, Van Der Straeten C. What is the learning curve associated with a hip resurfacing? Annals of Joint. 2019;4.
142. Nunley RM, Zhu J, Brooks PJ, Engh CA, Jr., Raterman SJ, Rogerson JS, et al. The learning curve for adopting hip resurfacing among hip specialists. Clin Orthop Relat Res. 2010;468(2):382-91.
143. Seyler TM, Lai LP, Sprinkle DI, Ward WG, Jinnah RH. Does computer-assisted surgery improve accuracy and decrease the learning curve in hip resurfacing? A radiographic analysis. J Bone Joint Surg Am. 2008;90 Suppl 3:71-80.
144. Koper M, Reijman M, van Es E, Waarsing J, Koot H, Keizer S, et al. No added value for Computer-Assisted surgery to improve femoral component positioning and Patient Reported Outcomes in Hip Resurfacing Arthroplasty; a multi-center randomized controlled trial. BMC Musculoskeletal Disorders. 2019;20:1-9.
145. Matharu GS, Daniel J, Ziaee H, McMinn DJW. Failure of a Novel Ceramic-on-Ceramic Hip Resurfacing Prosthesis. The Journal of Arthroplasty. 2015;30(3):416-8.
146. Atrey A, Waite J, Hart A, Meswania J, Morison Z, Carr R, et al. Failure of a Ceramic-on-Ceramic Hip Resurfacing Due to Metallosis: A Case Report. JBJS Case Connect. 2014;4(1):e28.
147. Trentani C, Vaccarino F. The Paltrinieri-Trentani hip joint resurface arthroplasty. Clin Orthop Relat Res. 1978(134):36-40.
148. Kim WC, Grogan T, Amstutz HC, Dorey F. Survivorship comparison of THARIES and conventional hip arthroplasty in patients younger than 40 years old. Clin Orthop Relat Res. 1987(214):269-77.
149. Treacy RBC, Holland JP, Daniel J, Ziaee H, McMinn DJW. Preliminary report of clinical experience with metal-on-highly-crosslinked-polyethylene hip resurfacing. Bone Joint Res. 2019;8(10):443-50.
150. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg. 2011;128(1):305-10.
151. OCEBM Working Group. The Oxford Levels of Evidence 2. Oxford Centre for Evidence Based Medicine, http://wwwcebmnet/indexaspx?o=5653. 2011.
152. Wright JG, Swiontkowski MF, Heckman JD. Introducing Levels of Evidence to The Journal. JBJS. 2003;85(1):1-3.
153. Middleton R, Price A. Levels of evidence in orthopaedics. Journal of ISAKOS. 2017;2(6):295-6.
154. Liddle AD. Failure of Unicompartmental Knee Replacement: University of Oxford; 2013.
155. Adashek JJ, Kurzrock R. Balancing clinical evidence in the context of a pandemic. Nature Biotechnology. 2021;39(3):270-4.
156. Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, et al. The role of machine learning in clinical research: transforming the future of evidence generation. Trials. 2021;22(1):1-15.
157. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. Clin Cancer Res. 2019;25(13):3753-8.
158. Wedam S, Fashoyin-Aje L, Bloomquist E, Tang S, Sridhara R, Goldberg KB, et al. FDA Approval Summary: Palbociclib for Male Patients with Metastatic Breast Cancer. Clinical Cancer Research. 2020;26(6):1208-12.
159. Paradis C. Bias in surgical research. Annals of surgery. 2008;248(2):180-8.
160. Campbell AJ, Bagley A, Van Heest A, James MA. Challenges of Randomized Controlled Surgical Trials. Orthopedic Clinics of North America. 2010;41(2):145-55.

161. Farrow L, Gardner WT, Ablett AD, Kutuzov V, Johnstone A. A review of trauma and orthopaedic randomised clinical trials published in high-impact general medical journals. European Journal of Orthopaedic Surgery & Traumatology. 2021:1-11.
162. Smith CS, Mollon B, Vannabouathong C, Fu JM, Sales B, Bhandari M, et al. An assessment of randomized controlled trial quality in the journal of bone & joint surgery: update from 2001 to 2013. JBJS. 2020;102(20):e116.
163. Realpe AX, Blackstone J, Griffin DR, Bing AJF, Karski M, Milner SA, et al. Barriers to recruitment to an orthopaedic randomized controlled trial comparing two surgical procedures for ankle arthritis : a qualitative study. Bone Jt Open. 2021;2(8):631-7.
164. Gazendam A, Ekhtiari S, Rubinger L, Bhandari M. Common errors in the design of orthopaedic trials: Has anything changed? Injury. 2023;54:S43-S5.
165. Kluzek S, Dean B, Wartolowska KA. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. BMJ Evidence-Based Medicine. 2022;27(3):153-5.
166. Browne JA, Springer B, Spindler KP. Optimizing Use of Large Databases in Joint Arthroplasty and Orthopaedics. JBJS. 2022;104(Suppl 3):28-32.
167. Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. Gastroenterol Hepatol Bed Bench. 2012;5(2):79-83.
168. Kahlert J, Gribsholt SB, Gammelager H, Dekkers OM, Luta G. Control of confounding in the analysis phase - an overview for clinicians. Clin Epidemiol. 2017;9:195-204.
169. The Swedish Arthroplasty Register. About the register. 2023 [Available from: https://sar.registercentrum.se/.
170. The National Joint Registry for England and Wales. About the NJR 2023 [Available from: https://www.njrcentre.org.uk/about-us/.
171. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol. 2015;12:14.
172. Gliklich RE, Leavy MB, Dreyer NA. Analysis, interpretation, and reporting of registry data to evaluate outcomes. Registries for Evaluating Patient Outcomes: A User's Guide [Internet] 4th edition: Agency for Healthcare Research and Quality (US); 2020.
173. Hansen VJ, Greene ME, Bragdon MA, Nebergall AK, Barr CJ, Huddleston J, et al. Registries collecting level-I through IV Data: institutional and multicenter use: AAOS exhibit selection. JBJS. 2014;96(18):e160.
174. Gomes LSM, Roos MV, Takata ET, Schuroff AA, Alves SD, Camisa A, et al. Advantages and limitations of national arthroplasty registries. The need for multicenter registries: the Rempro-SBQ☆. Revista Brasileira de Ortopedia. 2017;52:3-13.
175. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw. 2011;39(5):1-13.
176. Shao Z, Bi S. Patient satisfaction after total hip arthroplasty: Influencing factors. Frontiers in Surgery. 2023;9:1043508.
177. Lie SA, Fenstad AM, Lygre SHL, Kroken G, Dybvik E, Gjertsen JE, et al. Kaplan-Meier and Cox Regression Are Preferable for the Analysis of Time to Revision of Joint Arthroplasty: Thirty-One Years of Follow-up for Cemented and Uncemented THAs Inserted From 1987 to 2000 in the Norwegian Arthroplasty Register. JB JS Open Access. 2022;7(1).
178. Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. Health Expect. 2021;24(4):1015-24.

179. Ben-Shlomo Y, Blom A, Boulton C, Brittain R, Clark E, Dawson-Bowling S, et al. Patient Reported Outcome Measures (PROMs) in the NJR. The National Joint Registry 19th Annual Report 2022 [Internet]: National Joint Registry; 2022.
180. Sokas C, Hu F, Edelen M, Sisodia R, Pusic A, Cooper Z. A Review of PROM Implementation in Surgical Practice. Annals of Surgery. 2022;275(1):85-90.
181. Rolfson O, Kärrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011;93(7):867-75.
182. Group TE. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.
183. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). British journal of rheumatology. 1997;36(5):551-9.
184. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Medical care. 1996;34(3):220-33.
185. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Bmj. 1992;305(6846):160-4.
186. Boonstra AM, Schiphorst Preuper HR, Reneman MF, Posthumus JB, Stewart RE. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. Int J Rehabil Res. 2008;31(2):165-9.
187. Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain. 1983;16(1):87-101.
188. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 1993;2(3):169-80.
189. Zampelis V, Ornstein E, Franzén H, Atroshi I. A simple visual analog scale for pain is as responsive as the WOMAC, the SF-36, and the EQ-5D in measuring outcomes of revision hip arthroplasty. Acta Orthopaedica. 2014;85(2):128-32.
190. Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. Quality of Life Research. 1998;7:155-66.
191. Nilsdotter A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. Arthritis Care Res (Hoboken). 2011;63 Suppl 11:S200-7.
192. Wamper KE, Sierevelt IN, Poolman RW, Bhandari M, Haverkamp D. The Harris hip score: Do ceiling effects limit its usefulness in orthopedics? Acta Orthop. 2010;81(6):703-7.
193. Rolfson O, Dahlberg L, Malchau H, Garellick G. Variables determining outcome in total hip replacement surgery. The Journal of Bone & Joint Surgery British Volume. 2009;91(2):157-61.
194. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833-40.
195. Roos MKLSLEM. WOMAC Osteoarthritis Index: Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Scandinavian Journal of Rheumatology. 1999;28(4):210-5.

196. Whitehouse SL, Crawford RW, Learmonth ID. Validation for the reduced Western Ontario and McMaster Universities Osteoarthritis Index function scale. J Orthop Surg (Hong Kong). 2008;16(1):50-3.
197. Ritter MA, Albohm MJ. Overview: maintaining outcomes for total hip arthroplasty. The past, present, and future. Clin Orthop Relat Res. 1997(344):81-7.
198. Ostendorf M, Van Stel H, Buskens E, Schrijvers A, Marting L, Verbout A, et al. Patient-reported outcome in total hip replacement: a comparison of five instruments of health status. The Journal of Bone & Joint Surgery British Volume. 2004;86(6):801-8.
199. Dawson J, Fitzpatrick R. Questionnaire on the perceptions of patients about total hip replacement. The Journal of Bone & Joint Surgery British Volume. 1996;78(2):185-90.
200. Arden NK, Kiran A, Judge A, Biant LC, Javaid MK, Murray DW, et al. What is a good patient reported outcome after total hip replacement? Osteoarthritis and Cartilage. 2011;19(2):155-62.
201. Field RE, Cronin MD, Singh PJ. The Oxford hip scores for primary and revision hip replacement. The Journal of Bone & Joint Surgery British Volume. 2005;87-B(5):618-22.
202. Murray D, Fitzpatrick R, Rogers K, Pandit H, Beard D, Carr A, et al. The use of the Oxford hip and knee scores. The Journal of Bone & Joint Surgery British Volume. 2007;89(8):1010-4.
203. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. Journal of Clinical Epidemiology. 2015;68(1):73-9.
204. Sabah SA, Alvand A, Beard DJ, Price AJ. Minimal important changes and differences were estimated for Oxford hip and knee scores following primary and revision arthroplasty. Journal of Clinical Epidemiology. 2022;143:159-68.
205. Holmenlund C, Overgaard S, Bilberg R, Varnum C. Evaluation of the Oxford Hip Score: Does it still have content validity? Interviews of total hip arthroplasty patients. Health and Quality of Life Outcomes. 2021;19(1):237.
206. Sayers A, Whitehouse MR, Judge A, MacGregor AJ, Blom AW, Ben-Shlomo Y. Analysis of change in patient-reported outcome measures with floor and ceiling effects using the multilevel Tobit model: a simulation study and an example from a National Joint Register using body mass index and the Oxford Hip Score. BMJ Open. 2020;10(8):e033646.
207. Ingelsrud LH, Wilkinson JM, Overgaard S, Rolfson O, Hallstrom B, Navarro RA, et al. How do Patient-reported Outcome Scores in International Hip and Knee Arthroplasty Registries Compare? Clin Orthop Relat Res. 2022;480(10):1884-96.
208. Edwards TC, Guest B, Garner A, Logishetty K, Liddle AD, Cobb JP. The metabolic equivalent of task score. Bone & Joint Research. 2022;11(5):317-26.
209. Wiik AV, Brevadt M, Johal H, Logishetty K, Boughton O, Aqil A, et al. The loading patterns of a short femoral stem in total hip arthroplasty: gait analysis at increasing walking speeds and inclines. Journal of Orthopaedics and Traumatology. 2018;19(1):14.
210. Wiik AV, Aqil A, Al-Obaidi B, Brevadt M, Cobb JP. The impact of reducing the femoral stem length in total hip arthroplasty during gait. Archives of Orthopaedic and Trauma Surgery. 2021;141(11):1993-2000.
211. Aqil A, Drabu R, Bergmann JH, Masjedi M, Manning V, Andrews B, et al. The gait of patients with one resurfacing and one replacement hip: a single blinded controlled study. International Orthopaedics. 2013;37(5):795-801.
212. Teufl W, Taetz B, Miezal M, Lorenz M, Pietschmann J, Jöllenbeck T, et al. Towards an Inertial Sensor-Based Wearable Feedback System for Patients after Total Hip Arthroplasty: Validity and Applicability for Gait Classification with Gait Kinematics-Based Features. Sensors. 2019;19(22):5006.
213. Reininga IH, Stevens M, Wagenmakers R, Boerboom AL, Groothoff JW, Bulstra SK, et al. Comparison of gait in patients following a computer-navigated minimally

invasive anterior approach and a conventional posterolateral approach for total hip arthroplasty: A randomized controlled trial. Journal of Orthopaedic Research. 2013;31(2):288-94.
214. 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 musculoskeletal disorders. 2012;13(1):1-13.
215. Jolles B, Zangger P. Factors predisposing to dislocation after primary total hip arthroplasty: a multivariate analysis. The Journal of arthroplasty. 2002;17(3):282-8.
216. Karachalios T, Komnos G, Koutalos A. Total hip arthroplasty: survival and modes of failure. EFORT open reviews. 2018;3(5):232-9.
217. Parvizi J, Johnson BG, Rowland C, Ereth MH, Lewallen DG. Thirty-Day Mortality After Elective Total Hip Arthroplasty. JBJS. 2001;83(10):1524-8.
218. Evans JT, Blom AW, Timperley AJ, Dieppe P, Wilson MJ, Sayers A, et al. Factors associated with implant survival following total hip replacement surgery: A registry study of data from the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man. PLoS Med. 2020;17(8):e1003291.
219. Jämsen E, Puolakka T, Eskelinen A, Jäntti P, Kalliovalkama J, Nieminen J, et al. Predictors of mortality following primary hip and knee replacement in the aged. A single-center analysis of 1,998 primary hip and knee replacements for primary osteoarthritis. Acta Orthop. 2013;84(1):44-53.
220. Singh JA, Lewallen DG. Ninety-day mortality in patients undergoing elective total hip or total knee arthroplasty. J Arthroplasty. 2012;27(8):1417-22.e1.
221. Miller KA, Callaghan JJ, Goetz DD, Johnston RC. Early postoperative mortality following total hip arthroplasty in a community setting: a single surgeon experience. Iowa Orthop J. 2003;23:36-42.
222. Blom A, Pattison G, Whitehouse S, Taylor A, Bannister G. Early death following primary total hip arthroplasty: 1,727 procedures with mechanical thrombo-prophylaxis. Acta Orthop. 2006;77(3):347-50.
223. Turan O, Pan X, Kunze KN, Rullan PJ, Emara AK, Molloy RM, et al. 30-day to 10-year mortality rates following total hip arthroplasty: a meta-analysis of the last decade (2011–2021). HIP International. 2023:11207000231151235.
224. Ramiah RD, Ashmore AM, Whitley E, Bannister GC. Ten-year life expectancy after primary total hip replacement. The Journal of Bone & Joint Surgery British Volume. 2007;89-B(10):1299-302.
225. Harris IA, Hatton A, Pratt N, Lorimer M, Naylor JM, de Steiger R, et al. How Does Mortality Risk Change Over Time After Hip and Knee Arthroplasty? Clin Orthop Relat Res. 2019;477(6):1414-21.
226. Ebeling M, Rau R, Malmström H, Ahlbom A, Modig K. The rate by which mortality increase with age is the same for those who experienced chronic disease as for the general population. Age and Ageing. 2021;50(5):1633-40.
227. Deere K, Whitehouse MR, Kunutsor SK, Sayers A, Mason J, Blom AW. How long do revised and multiply revised hip replacements last? A retrospective observational study of the National Joint Registry. The Lancet Rheumatology. 2022;4(7):e468-e79.
228. 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. 2019;393(10172):647-54.
229. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. The effect of age on pain, function, and quality of life after total hip and knee arthroplasty. Arch Intern Med. 2001;161(3):454-60.

230. Clement ND, MacDonald D, Howie CR, Biant LC. The outcome of primary total hip and knee arthroplasty in patients aged 80 years or more. J Bone Joint Surg Br. 2011;93(9):1265-70.
231. Lawless BM, Greene M, Slover J, Kwon YM, Malchau H. Does age or bilateral disease influence the value of hip arthroplasty? Clin Orthop Relat Res. 2012;470(4):1073-8.
232. Murphy BPD, Dowsey MM, Spelman T, Choong PFM. What Is the Impact of Advancing Age on the Outcomes of Total Hip Arthroplasty? J Arthroplasty. 2018;33(4):1101-7.e1.
233. Nilsdotter AK, Lohmander LS. Age and waiting time as predictors of outcome after total hip replacement for osteoarthritis. Rheumatology. 2002;41(11):1261-7.
234. Joly DA, Ludwig T, Mahdavi S, Khong H, Piroozfar SG, Sharma R. Does Age Influence Patient-Reported Outcomes in Unilateral Primary Total Hip and Knee Arthroplasty? J Arthroplasty. 2020;35(7):1800-5.
235. Götz JS, Benditz A, Reinhard J, Schindler M, Zeman F, Grifka J, et al. Influence of Anxiety/Depression, Age, Gender and ASA on 1-Year Follow-Up Outcomes Following Total Hip and Knee Arthroplasty in 5447 Patients. J Clin Med. 2021;10(14).
236. Marques CJ, Bohlen K, Lampe F. Participation in a Preoperative Patient Education Session Is a Significant Predictor of Better WOMAC Total Index Score and Higher EQ-5D-5L Health Status Index 1 Year After Total Knee and Hip Arthroplasties: A Retrospective Observational Study. Am J Phys Med Rehabil. 2021;100(10):972-7.
237. Navas L, Faller J, Schmidt S, Streit M, Hauschild M, Zimmerer A. Sports Activity and Patient-Related Outcomes after Cementless Total Hip Arthroplasty in Patients Younger than 40 Years. J Clin Med. 2021;10(20).
238. Okafor L, Chen AF. Patient satisfaction and total hip arthroplasty: a review. Arthroplasty. 2019;1(1):6.
239. Falez F, Mavrogenis A, Scarlat MM. Outcome scores after hip surgery in young adults: an editorial approach. Int Orthop. 2022;46(8):1675-9.
240. Clement ND, Smith KM, Baron YJ, McColm H, Deehan DJ, Holland J. Increasing age does not influence hip-specific functional outcome or health-related quality of life following total hip arthroplasty : a five-year prospective cohort study. Bone Jt Open. 2022;3(9):692-700.
241. Berstock JR, Beswick AD, Lenguerrand E, Whitehouse MR, Blom AW. Mortality after total hip replacement surgery. Bone & Joint Research. 2014;3(6):175-82.
242. Gruberg L, Weissman NJ, Waksman R, Fuchs S, Deible R, Pinnow EE, et al. The impact of obesity on the short-term andlong-term outcomes after percutaneous coronary intervention: the obesity paradox? Journal of the American College of Cardiology. 2002;39(4):578-84.
243. Halawi MJ, Brigati D, Messner W, Brooks PJ. Total hip arthroplasty in patients 55 years or younger: risk factors for poor midterm outcomes. Journal of clinical orthopaedics and trauma. 2018;9(2):103-6.
244. Bülow E, Rolfson O, Cnudde P, Rogmark C, Garellick G, Nemes S. Comorbidity does not predict long-term mortality after total hip arthroplasty. Acta Orthopaedica. 2017;88(5):472-7.
245. Bozic KJ, Lau E, Kurtz S, Ong K, Rubash H, Vail TP, et al. Patient-related risk factors for periprosthetic joint infection and postoperative mortality following total hip arthroplasty in Medicare patients. J Bone Joint Surg Am. 2012;94(9):794-800.
246. 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;342:d1165.

247. Huddleston JI, Wang Y, Uquillas C, Herndon JH, Maloney WJ. Age and obesity are risk factors for adverse events after total hip arthroplasty. Clin Orthop Relat Res. 2012;470(2):490-6.
248. Tohidi M, Brogly SB, Lajkosz K, Harrison MM, Campbell AR, VanDenKerkhof E, et al. Ten-year risk of complication and mortality after total hip arthroplasty in morbidly obese patients: a population study. Can J Surg. 2019;62(6):442-9.
249. Jeschke E, Citak M, Günster C, Halder AM, Heller K-D, Malzahn J, et al. Obesity increases the risk of postoperative complications and revision rates following primary total hip arthroplasty: an analysis of 131,576 total hip arthroplasty cases. The Journal of arthroplasty. 2018;33(7):2287-92. e1.
250. Woo SH, Cha DH, Park E-C, Kim SJ. The association of under-weight and obesity with mortality after hip arthroplasty. Age and ageing. 2019;48(1):94-100.
251. Katakam A, Melnic CM, Bragdon CR, Sauder N, Collins AK, Bedair HS. Low body mass index is a predictor for mortality and increased length of stay following total joint arthroplasty. The Journal of Arthroplasty. 2021;36(1):72-7.
252. Buechele G, Guenther K-P, Brenner H, Puhl W, Stürmer T, Rothenbacher D, et al. Osteoarthritis-patterns, cardio-metabolic risk factors and risk of all-cause mortality: 20 years follow-up in patients after hip or knee replacement. Scientific reports. 2018;8(1):5253.
253. Cannata F, Laudisio A, Ambrosio L, Vadalà G, Russo F, Zampogna B, et al. The Association of Body Mass Index with Surgical Time Is Mediated by Comorbidity in Patients Undergoing Total Hip Arthroplasty. J Clin Med. 2021;10(23).
254. Bowditch MG, Villar RN. Do obese patients bleed more? A prospective study of blood loss at total hip replacement. Ann R Coll Surg Engl. 1999;81(3):198-200.
255. Liu W, Wahafu T, Cheng M, Cheng T, Zhang Y, Zhang X. The influence of obesity on primary total hip arthroplasty outcomes: A meta-analysis of prospective cohort studies. Orthopaedics & Traumatology: Surgery & Research. 2015;101(3):289-96.
256. Onggo JR, Onggo JD, de Steiger R, Hau R. Greater risks of complications, infections, and revisions in the obese versus non-obese total hip arthroplasty population of 2,190,824 patients: a meta-analysis and systematic review. Osteoarthritis Cartilage. 2020;28(1):31-44.
257. Goodnough LH, Finlay AK, Huddleston JI, Goodman SB, Maloney WJ, Amanatullah DF. Obesity Is Independently Associated With Early Aseptic Loosening in Primary Total Hip Arthroplasty. The Journal of Arthroplasty. 2018;33(3):882-6.
258. Jung P, Morris AJ, Zhu M, Roberts SA, Frampton C, Young SW. BMI is a key risk factor for early periprosthetic joint infection following total hip and knee arthroplasty. N Z Med J. 2017;130(1461):24-34.
259. Kennedy JW, Young D, Meek DRM, Patil SR. Obesity is associated with higher complication rates in revision total hip arthroplasty. J Orthop. 2018;15(1):70-2.
260. 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;90(3):220-5.
261. McDonald CL, Alsoof D, Johnson KG, Kuczmarski A, Lemme NJ, Testa EJ, et al. Underweight Patients are at Increased Risk for Complications following Total Hip Arthroplasty. The Journal of Arthroplasty. 2023;38(8):1559-64.e1.
262. Zusmanovich M, Kester B, Feng J, Schwarzkopf R. Postoperative complications in underweight patients undergoing total hip arthroplasty: A comparative analysis to normal weight patients. J Orthop. 2018;15(2):345-8.
263. Sayeed Z, Anoushiravani AA, Chambers MC, Gilbert TJ, Scaife SL, El-Othmani MM, et al. Comparing in-hospital total joint arthroplasty outcomes and resource

consumption among underweight and morbidly obese patients. The Journal of arthroplasty. 2016;31(10):2085-90.
264. Andrew JG, Palan J, Kurup HV, Gibson P, Murray DW, Beard DJ. Obesity in total hip replacement. J Bone Joint Surg Br. 2008;90(4):424-9.
265. Lash H, Hooper G, Hooper N, Frampton C. Should a patients BMI status be used to restrict access to total hip and knee arthroplasty? functional outcomes of arthroplasty relative to BMI-single centre retrospective review. The open orthopaedics journal. 2013;7:594.
266. Chee YH, Teoh KH, Sabnis BM, Ballantyne JA, Brenkel IJ. Total hip replacement in morbidly obese patients with osteoarthritis: results of a prospectively matched study. J Bone Joint Surg Br. 2010;92(8):1066-71.
267. Michalka PKR, Khan RJK, Scaddan MC, Haebich S, Chirodian N, Wimhurst JA. The Influence of Obesity on Early Outcomes in Primary Hip Arthroplasty. The Journal of Arthroplasty. 2012;27(3):391-6.
268. Haebich SJ, Mark P, Khan RJK, Fick DP, Brownlie C, Wimhurst JA. The Influence of Obesity on Hip Pain, Function, and Satisfaction 10 Years Following Total Hip Arthroplasty. The Journal of Arthroplasty. 2020;35(3):818-23.
269. McLaughlin JR, Lee KR. The outcome of total hip replacement in obese and non-obese patients at 10- to 18-years. The Journal of Bone & Joint Surgery British Volume. 2006;88-B(10):1286-92.
270. Tai SM, Imbuldeniya AM, Munir S, Walter WL, Walter WK, Zicat BA. The Effect of Obesity on the Clinical, Functional and Radiological Outcome of Cementless Total Hip Replacement: A Case-Matched Study With a Minimum 10-Year Follow-Up. The Journal of Arthroplasty. 2014;29(9):1758-62.
271. Robertson TS, Callary SA, Costi K, Clothier RJ, Venugopal K, Rickman M. The effect of weight compared to BMI on patient reported outcomes at long term follow up of primary total hip arthroplasty. Journal of Orthopaedics. 2023;41:14-22.
272. Aynardi M, Jacovides CL, Huang R, Mortazavi SM, Parvizi J. Risk factors for early mortality following modern total hip arthroplasty. J Arthroplasty. 2013;28(3):517-20.
273. Johnson RL, Habermann EB, Johnson MQ, Abdel MP, Chamberlain AM, Mantilla CB. How is surgical risk best assessed? A cohort comparison of measures in Total joint arthroplasty. The Journal of Arthroplasty. 2021;36(3):851-6. e3.
274. McConaghy KM, Orr MN, Grits D, Emara AK, Molloy RM, Piuzzi NS. What is the 30-day mortality burden after elective total hip arthroplasty? an analysis of 194,062 patients. The Journal of Arthroplasty. 2021;36(10):3513-8. e2.
275. Comba F, Alonso Hidalgo I, Buttaro M, Piccaluga F. Risk Factor Analysis for 30-Day Mortality After Primary THA in a Single Institution. Hss j. 2012;8(2):111-5.
276. Memtsoudis SG, Pumberger M, Ma Y, Chiu YL, Fritsch G, Gerner P, et al. Epidemiology and risk factors for perioperative mortality after total hip and knee arthroplasty. J Orthop Res. 2012;30(11):1811-21.
277. Arias-de la Torre J, Smith K, Dregan A, Valderas JM, Evans JP, Prieto-Alhambra D, et al. Impact of comorbidity on the short- and medium-term risk of revision in total hip and knee arthroplasty. BMC Musculoskelet Disord. 2020;21(1):447.
278. Podmore B, Hutchings A, Meulen Jvd, Aggarwal A, Konan S. Impact of comorbid conditions on outcomes of hip and knee replacement surgery: a systematic review and meta-analysis. BMJ Open. 2018;8(7):e021784.
279. Loth FL, Giesinger JM, Giesinger K, MacDonald DJ, Simpson A, Howie CR, et al. Impact of Comorbidities on Outcome After Total Hip Arthroplasty. J Arthroplasty. 2017;32(9):2755-61.

280. Greene ME, Rolfson O, Gordon M, Garellick G, Nemes S. Standard Comorbidity Measures Do Not Predict Patient-reported Outcomes 1 Year After Total Hip Arthroplasty. Clin Orthop Relat Res. 2015;473(11):3370-9.
281. Lan P, Chen X, Fang Z, Zhang J, Liu S, Liu Y. Effects of Comorbidities on Pain and Function After Total Hip Arthroplasty. Front Surg. 2022;9:829303.
282. Patel AP, Gronbeck C, Chambers M, Harrington MA, Halawi MJ. Gender and Total Joint Arthroplasty: Variable Outcomes by Procedure Type. Arthroplast Today. 2020;6(3):517-20.
283. Chen A, Paxton L, Zheng X, Peat R, Mao J, Liebeskind A, et al. Association of Sex With Risk of 2-Year Revision Among Patients Undergoing Total Hip Arthroplasty. JAMA Netw Open. 2021;4(6):e2110687.
284. Kostamo T, Bourne RB, Whittaker JP, McCalden RW, MacDonald SJ. No difference in gender-specific hip replacement outcomes. Clin Orthop Relat Res. 2009;467(1):135-40.
285. Ackerman IN, Busija L, Lorimer M, de Steiger R, Graves SE. Lifetime Risk of Revision Hip Replacement Surgery in Australia Remains Low: A Population-Level Analysis Using National Registry Data. J Bone Joint Surg Am. 2021;103(5):389-96.
286. Zhang B, Rao S, Mekkawy KL, Rahman R, Sarfraz A, Hollifield L, et al. Risk factors for pain after total hip arthroplasty: a systematic review. Arthroplasty. 2023;5(1):19.
287. Mannion AF, Impellizzeri FM, Naal FD, Leunig M. Women demonstrate more pain and worse function before THA but comparable results 12 months after surgery. Clin Orthop Relat Res. 2015;473(12):3849-57.
288. Peters RM, van Steenbergen LN, Stewart RE, Stevens M, Rijk PC, Bulstra SK, et al. Which patients improve most after total hip arthroplasty? Influence of patient characteristics on patient-reported outcome measures of 22,357 total hip arthroplasties in the Dutch Arthroplasty Register. HIP International. 2021;31(5):593-602.
289. CRAWFORD RW, MURRAY DW. Total hip replacement: indications for surgery and risk factors for failure. Annals of the Rheumatic Diseases. 1997;56(8):455-7.
290. Moralidou M, Di Laura A, Henckel J, Hothi H, Hart AJ. Three-dimensional pre-operative planning of primary hip arthroplasty: a systematic literature review. EFORT Open Rev. 2020;5(12):845-55.
291. Mufarrih SH, Ghani MOA, Martins RS, Qureshi NQ, Mufarrih SA, Malik AT, et al. Effect of hospital volume on outcomes of total hip arthroplasty: a systematic review and meta-analysis. J Orthop Surg Res. 2019;14(1):468.
292. Jolbäck P, Rolfson O, Mohaddes M, Nemes S, Kärrholm J, Garellick G, et al. Does surgeon experience affect patient-reported outcomes 1 year after primary total hip arthroplasty? Acta Orthop. 2018;89(3):265-71.
293. Blom AW, Hunt LP, Matharu GS, Reed MR, Whitehouse MR. The effect of surgical approach in total hip replacement on outcomes: an analysis of 723,904 elective operations from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. BMC Medicine. 2020;18(1):242.
294. Miller LL, Prieto-Alhambra D, Trela-Larsen L, Wilkinson JM, Clark EM, Blom AW, et al. Revision and 90-day mortality following hip arthroplasty in patients with inflammatory arthritis and ankylosing spondylitis enrolled in the National Joint Registry for England and Wales. Hip Int. 2022;32(3):371-8.
295. Zhang Y, Chu SS, Liu K, Huang Q, Wang Y. Outcomes in patients with rheumatoid versus osteoarthritis for total hip arthroplasty: A meta-analysis and systematic review. Semin Arthritis Rheum. 2022;56:152061.
296. Le Manach Y, Collins G, Bhandari M, Bessissow A, Boddaert J, Khiami F, et al. Outcomes After Hip Fracture Surgery Compared With Elective Total Hip Replacement. JAMA. 2015;314(11):1159-66.

297. Szczesiul J, Bielecki M. A Review of Total Hip Arthroplasty Comparison in FNF and OA Patients. Adv Orthop. 2021;2021:5563500.
298. Siddiqi A, White PB, Sloan M, Fox D, Piuzzi NS, Sankar WN, et al. Total Hip Arthroplasty for Developmental Dysplasia of Hip vs Osteoarthritis: A Propensity Matched Pair Analysis. Arthroplast Today. 2020;6(3):607-11.e1.
299. Cram P, Vaughan-Sarrazin MS, Wolf B, Katz JN, Rosenthal GE. A comparison of total hip and knee replacement in specialty and general hospitals. J Bone Joint Surg Am. 2007;89(8):1675-84.
300. Shervin N, Rubash HE, Katz JN. Orthopaedic procedure volume and patient outcomes: a systematic literature review. Clin Orthop Relat Res. 2007;457:35-41.
301. 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 Joint Surg Am. 2008;90(11):2446-51.
302. Oakley CT, Arraut J, Lygrisse K, Schwarzkopf R, Slover JD, Rozell JC. The effect of surgeon and hospital volume on total hip arthroplasty patient-reported outcome measures: an American joint replacement registry study. Journal of the American Academy of Orthopaedic Surgeons. 2023;31(4):205-11.
303. D'Apuzzo MR, Higgins MD, Novicoff WM, Browne JA. Hospital Volume as a Source of Variation for Major Complications and Early In-Hospital Mortality After Total Joint Arthroplasty. Arthroplast Today. 2022;16:53-6.
304. Bozic KJ, Maselli J, Pekow PS, Lindenauer PK, Vail TP, Auerbach AD. The influence of procedure volumes and standardization of care on quality and efficiency in total joint replacement surgery. J Bone Joint Surg Am. 2010;92(16):2643-52.
305. Singh JA, Kwoh CK, Boudreau RM, Lee GC, Ibrahim SA. Hospital volume and surgical outcomes after elective hip/knee arthroplasty: a risk-adjusted analysis of a large regional database. Arthritis Rheum. 2011;63(8):2531-9.
306. Doro C, Dimick J, Wainess R, Upchurch G, Urquhart A. Hospital Volume and Inpatient Mortality Outcomes of Total Hip Arthroplasty in the United States. The Journal of Arthroplasty. 2006;21(6, Supplement):10-6.
307. Ravi B, Jenkinson R, Austin PC, Croxford R, Wasserstein D, Escott B, et al. Relation between surgeon volume and risk of complications after total hip arthroplasty: propensity score matched cohort study. Bmj. 2014;348:g3284.
308. Sayers A, Steele F, Whitehouse MR, Price A, Ben-Shlomo Y, Blom AW. Association between surgical volume and failure of primary total hip replacement in England and Wales: findings from a prospective national joint replacement register. BMJ Open. 2020;10(9):e033045.
309. Yan L, Ge L, Dong S, Saluja K, Li D, Reddy KS, et al. Evaluation of Comparative Efficacy and Safety of Surgical Approaches for Total Hip Arthroplasty: A Systematic Review and Network Meta-analysis. JAMA Netw Open. 2023;6(1):e2253942.
310. Koster M, Luzier AD, Temmerman OP, Vos SJ, Benner JL. How do dislocation rates differ between different approaches to total hip arthroplasty? A systematic review and meta-analysis. Journal of Orthopaedics, Trauma and Rehabilitation. 2023;30(1):22104917221147688.
311. Pincus D, Jenkinson R, Paterson M, Leroux T, Ravi B. Association Between Surgical Approach and Major Surgical Complications in Patients Undergoing Total Hip Arthroplasty. Jama. 2020;323(11):1070-6.
312. Chatfield C. Reducing variations in care. Bmj. 2017;359:j5574.
313. Briggs T. A national review of adult elective orthopaedic services in England. Getting it right first time. 2015;7.

314. Sharkey PF, Sethuraman V, Hozack WJ, Rothman RH, Stiehl JB. Factors influencing choice of implants in total hip arthroplasty and total knee arthroplasty: perspectives of surgeons and patients. J Arthroplasty. 1999;14(3):281-7.
315. Gardezi M, Ottesen TD, Tyagi V, Sherman JJZ, Grauer JN, Rubin LE. Arthroplasty implants and materials: Cost awareness and value perception. PLoS One. 2021;16(7):e0255061.
316. Park CW, Lim SJ, Park YS. Modular Stems: Advantages and Current Role in Primary Total Hip Arthroplasty. Hip Pelvis. 2018;30(3):147-55.
317. Solarino G, Vicenti G, Carrozzo M, Ottaviani G, Moretti B, Zagra L. Modular neck stems in total hip arthroplasty: current concepts. EFORT Open Reviews. 2021;6(9):751-8.
318. Manfreda F, Bufi E, Talesa GR, Florio EF, Caraffa A. Is there a real difference between modular stems and monoblock implants in THA? A revision and comparison of tribological, clinical and radiological outcomes. Lo Scalpello-Journal. 2021;35:130-8.
319. Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. Lancet. 2012;379(9822):1199-204.
320. Matar HE, Platt SR, Board TN, Porter ML. Overview of Randomized Controlled Trials in Primary Total Hip Arthroplasty (34,020 Patients): What Have We Learnt? J Am Acad Orthop Surg Glob Res Rev. 2020;4(8):e20.00120.
321. Zheng C, Xu J, Wu L, Wu Y, Liu Y, Shen B. Comparisons of different bearing surfaces in cementless total hip arthroplasty: a systematic review and Bayesian network analysis. The Journal of Arthroplasty. 2022.
322. López-López JA, Humphriss RL, Beswick AD, Thom HHZ, Hunt LP, Burston A, et al. Choice of implant combinations in total hip replacement: systematic review and network meta-analysis. Bmj. 2017;359:j4651.
323. Hanna SA, Sewell MD, Sri-Ram K, Miles J, Aston WJS, Pollock RC, et al. The Effect of Femoral Head Size on Functional Outcome in Primary Total Hip Arthroplasty: A Single-Blinded Randomised Controlled Trial. HIP International. 2012;22(6):592-7.
324. Howie DW, Holubowycz OT, Middleton R, Group tLAS. Large Femoral Heads Decrease the Incidence of Dislocation After Total Hip Arthroplasty: A Randomized Controlled Trial. JBJS. 2012;94(12):1095-102.
325. Tsikandylakis G, Mohaddes M, Cnudde P, Eskelinen A, Kärrholm J, Rolfson O. Head size in primary total hip arthroplasty. EFORT Open Rev. 2018;3(5):225-31.
326. Girard J. Femoral head diameter considerations for primary total hip arthroplasty. Orthop Traumatol Surg Res. 2015;101(1 Suppl):S25-9.
327. Young EY, Gebhart J, Cooperman D, Ahn NU. Are the left and right proximal femurs symmetric? Clin Orthop Relat Res. 2013;471(5):1593-601.
328. Brooks PJ, Samuel LT, Levin JM, Sultan AA, Khlopas A, Brigati D, et al. Mortality after hip resurfacing versus total hip arthroplasty in young patients: a single surgeon experience. Ann Transl Med. 2019;7(4):77.
329. Palazzuolo M, Bensa A, Bauer S, Blakeney WG, Filardo G, Riegger M. Resurfacing Hip Arthroplasty Is a Safe and Effective Alternative to Total Hip Arthroplasty in Young Patients: A Systematic Review and Meta-Analysis. J Clin Med. 2023;12(6).
330. Costa ML, Achten J, Foguet P, Parsons NR. Comparison of hip function and quality of life of total hip arthroplasty and resurfacing arthroplasty in the treatment of young patients with arthritis of the hip joint at 5 years. BMJ Open. 2018;8(3):e018849.
331. Garbuz DS, Tanzer M, Greidanus NV, Masri BA, Duncan CP. The John Charnley Award: Metal-on-metal hip resurfacing versus large-diameter head metal-on-metal total hip arthroplasty: a randomized clinical trial. Clin Orthop Relat Res. 2010;468(2):318-25.

332. Oxblom A, Hedlund H, Nemes S, Brismar H, Felländer-Tsai L, Rolfson O. Patient-reported outcomes in hip resurfacing versus conventional total hip arthroplasty: a register-based matched cohort study of 726 patients. Acta Orthop. 2019;90(4):318-23.
333. Kumar P, Ksheersagar V, Aggarwal S, Jindal K, Dadra A, Kumar V, et al. Complications and mid to long term outcomes for hip resurfacing versus total hip replacement: a systematic review and meta-analysis. Eur J Orthop Surg Traumatol. 2023;33(5):1495-504.
334. Magan A, Wignadasan W, Kayani B, Radhakrishnan G, Ronca F, Haddad FS. A meta-analysis assessing time for return to sport following hip resurfacing. Arch Orthop Trauma Surg. 2023;143(6):3575-85.
335. Hellman MD, Ford MC, Barrack RL. Is there evidence to support an indication for surface replacement arthroplasty?: a systematic review. Bone Joint J. 2019;101-b(1_Supple_A):32-40.
336. Satalich JR, Lombardo DJ, Newman S, Golladay GJ, Patel NK. Cementation in total hip arthroplasty: history, principles, and technique. EFORT Open Rev. 2022;7(11):747-57.
337. McKee GK, Watson-Farrar J. Replacement of arthritic hips by the McKee-Farrar prosthesis. J Bone Joint Surg Br. 1966;48(2):245-59.
338. Wroblewski BM, Siney PD, Fleming PA. Charnley low-frictional torque arthroplasty in patients under the age of 51 years. Follow-up to 33 years. J Bone Joint Surg Br. 2002;84(4):540-3.
339. Sutherland CJ, Wilde AH, Borden LS, Marks KE. A ten-year follow-up of one hundred consecutive Müller curved-stem total hip-replacement arthroplasties. J Bone Joint Surg Am. 1982;64(7):970-82.
340. Yamada H, Yoshihara Y, Henmi O, Morita M, Shiromoto Y, Kawano T, et al. Cementless total hip replacement: past, present, and future. J Orthop Sci. 2009;14(2):228-41.
341. Haas SS, Brauer GM, Dickson G. A characterization of polymethylmethacrylate bone cement. J Bone Joint Surg Am. 1975;57(3):380-91.
342. Erivan R, Villatte G, Khelif YR, Pereira B, Galvin M, Descamps S, et al. The Müller self-locking cemented total hip prosthesis with polyethylene liner: After twenty years, what did they become? Int Orthop. 2017;41(1):47-54.
343. Kawamoto K, Hasegawa Y, Iwase T, Iwasada S, Kanamono T, Iwata H. Failed cementless total hip arthroplasty for osteoarthrosis due to hip dysplasia. A minimum five-year follow-up study. Bull Hosp Jt Dis. 1998;57(3):130-5.
344. Kazimierczak P, Przekora A. Osteoconductive and Osteoinductive Surface Modifications of Biomaterials for Bone Regeneration: A Concise Review. Coatings. 2020;10(10):971.
345. Hailer NP, Garellick G, Kärrholm J. Uncemented and cemented primary total hip arthroplasty in the Swedish Hip Arthroplasty Register. Acta Orthop. 2010;81(1):34-41.
346. Chen XT, Christ AB, Chung BC, Ton A, Ballatori AM, Shahrestani S, et al. Cemented versus Cementless Femoral Fixation for Elective Primary Total Hip Arthroplasty: A Nationwide Analysis of Short-Term Complication and Readmission Rates. Journal of Clinical Medicine. 2023;12(12):3945.
347. Zhang C, Yan CH, Zhang W. Cemented or cementless fixation for primary hip arthroplasty—evidence from The International Joint Replacement Registries. Annals of Joint. 2017;2(10).
348. Garland A, Gordon M, Garellick G, Kärrholm J, Sköldenberg O, Hailer N. Risk of early mortality after cemented compared with cementless total hip arthroplasty: a nationwide matched cohort study. The Bone & Joint Journal. 2017;99(1):37-43.
349. Corten K, Bourne RB, Charron KD, Au K, Rorabeck CH. What Works Best, a Cemented or Cementless Primary Total Hip Arthroplasty?: Minimum 17-year Followup

of a Randomized Controlled Trial. Clinical Orthopaedics and Related Research®. 2011;469(1):209-17.
350. Konan S, Abdel MP, Haddad FS. Cemented versus uncemented hip implant fixation. Bone & Joint Research. 2019;8(12):604-7.
351. Cao Y, Fang X, Ottosson J, Näslund E, Stenberg E. A comparative study of machine learning algorithms in predicting severe complications after bariatric surgery. Journal of Clinical Medicine. 2019;8(5):668.
352. Lee H-C, Yoon H-K, Nam K, Cho YJ, Kim TK, Kim WH, et al. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. Journal of clinical medicine. 2018;7(10):322.
353. Luo C, Zhu Y, Zhu Z, Li R, Chen G, Wang Z. A machine learning-based risk stratification tool for in-hospital mortality of intensive care unit patients with heart failure. Journal of Translational Medicine. 2022;20(1):136.
354. Oermann EK, Rubinsteyn A, Ding D, Mascitelli J, Starke RM, Bederson JB, et al. Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. Scientific reports. 2016;6(1):21161.
355. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach. Academic emergency medicine. 2016;23(3):269-78.
356. Liu W, Laranjo L, Klimis H, Chiang J, Yue J, Marschner S, et al. Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: a systematic review and meta-analysis. European Heart Journal-Quality of Care and Clinical Outcomes. 2023:qcad017.
357. Herm L-V, Heinrich K, Wanner J, Janiesch C. Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. International Journal of Information Management. 2023;69:102538.
358. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artificial Intelligence in healthcare: Elsevier; 2020. p. 25-60.
359. NJR Feedback Services 2023 [Available from: https://reports.njrcentre.org.uk/NJR-Feedback-Services.
360. Hospital Episode Statistics (HES) 2023 [Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics.
361. Bayliss LE, Culliford D, Monk AP, Glyn-Jones S, Prieto-Alhambra D, Judge A, 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. The Lancet. 2017;389(10077):1424-30.
362. Health SoSf. High quality care for all: NHS next stage review final report: The Stationery Office; 2008.
363. NHS England. National Patient Reported Outcome Measures (PROMs) Programme Consultation Report.; 2017.
364. Callaghan JJ, Dysart SH, Savory CF, Hopkinson WJ. Assessing the results of hip replacement. A comparison of five different rating systems. The Journal of Bone and Joint Surgery British volume. 1990;72(6):1008-9.
365. Group TE. EuroQol-a new facility for the measurement of health-related quality of life. Health policy. 1990;16(3):199-208.
366. Ackerman IN, Bohensky MA, Zomer E, Tacey M, Gorelik A, Brand CA, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. BMC musculoskeletal disorders. 2019;20:1-10.

367. Bloembergen CH, van de Graaf VA, Virgile A, Willigenburg NW, Noorduyn JC, Saris DB, et al. Infographic. Can even experienced orthopaedic surgeons predict who will benefit from surgery when patients present with degenerative meniscal tears? A survey of 194 orthopaedic surgeons who made 3880 predictions. British Journal of Sports Medicine. 2020;54(9):556-7.
368. Oosterhoff JH, Doornberg JN. Artificial intelligence in orthopaedics: false hope or not? A narrative review along the line of Gartner's hype cycle. EFORT open reviews. 2020;5(10):593-603.
369. Chapot A, Zambelli P-Y, Merckaert SR. Functional and Patient-related Outcomes of Total Hip Arthroplasty in Patients Younger Than 20 Years. Arthroplasty Today. 2023;20:101100.
370. Wade R, Shah KA. Functional and radiological outcome of uncemented total hip arthroplasty in young adults-5 year follow-upollow-up. Journal of Orthopaedics. 2020;18:237-9.
371. Zeitlinger L, Gemayel A, Whitlock P, Sorger J. Implant Survival and Clinical Outcomes of Total Hip Arthroplasty in Adolescent and Young Adult Patients. The Journal of Hip Surgery. 2021;5(02):062-9.
372. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. Otolaryngology—Head and Neck Surgery. 2010;143(3):331-6.
373. van Lingen CP, Zagra LM, Ettema HB, Verheyen CC. Sequelae of large-head metal-on-metal hip arthroplasties: Current status and future prospects. EFORT Open Rev. 2016;1(10):345-53.
374. Miguel-Hurtado O, Guest R, Stevenage SV, Neil GJ, Black S. Comparing machine learning classifiers and linear/logistic regression to explore the relationship between hand dimensions and demographic characteristics. PloS one. 2016;11(11):e0165521.
375. Seligman B, Tuljapurkar S, Rehkopf D. Machine learning approaches to the social determinants of health in the health and retirement study. SSM-population health. 2018;4:95-9.
376. Singal AG, Mukherjee A, Elmunzer BJ, Higgins PD, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. The American journal of gastroenterology. 2013;108(11):1723.
377. Garland A, Bülow E, Lenguerrand E, Blom A, Wilkinson M, Sayers A, et al. Prediction of 90-day mortality after total hip arthroplasty. Bone Joint J. 2021;103-b(3):469-78.
378. Trela-Larsen L, Kroken G, Bartz-Johannessen C, Sayers A, Aram P, McCloskey E, et al. Personalized estimation of one-year mortality risk after elective hip or knee arthroplasty for osteoarthritis. Bone & Joint Research. 2020;9(11):808-20.
379. Fassihi SC, Mathur A, Best MJ, Chen AZ, Gu A, Quan T, et al. Neural network prediction of 30-day mortality following primary total hip arthroplasty. J Orthop. 2021;28:91-5.
380. Klemt C, Laurencin S, Alpaugh K, Tirumala V, Barghi A, Yeo I, et al. The utility of machine learning algorithms for the prediction of early revision surgery after primary total hip arthroplasty. Journal of the American Academy of Orthopaedic Surgeons. 2022;30(11):513-22.
381. Klemt C, Cohen-Levy WB, Robinson MG, Burns JC, Alpaugh K, Yeo I, et al. Can machine learning models predict failure of revision total hip arthroplasty? Archives of Orthopaedic and Trauma Surgery. 2023;143(6):2805-12.

382. Turan O, Pan X, Kunze KN, Rullan PJ, Emara AK, Molloy RM, et al. 30-day to 10-year mortality rates following total hip arthroplasty: a meta-analysis of the last decade (2011–2021). HIP International. 2023;0(0):11207000231151235.
383. Bergh C, Fenstad AM, Furnes O, Garellick G, Havelin LI, Overgaard S, et al. Increased risk of revision in patients with non-traumatic femoral head necrosis. Acta Orthop. 2014;85(1):11-7.
384. Mont MA, Hungerford DS. Non-traumatic avascular necrosis of the femoral head. JBJS. 1995;77(3):459-74.
385. Cornell CN, Salvati EA, Pellicci PM. Long-term follow-up of total hip replacement in patients with osteonecrosis. The Orthopedic clinics of North America. 1985;16(4):757-69.
386. Kim Y-H, Choi Y, Kim J-S. Cementless total hip arthroplasty with ceramic-on-ceramic bearing in patients younger than 45 years with femoral-head osteonecrosis. International orthopaedics. 2010;34:1123-7.
387. Johannson HR, Zywiel MG, Marker DR, Jones LC, McGrath MS, Mont MA. Osteonecrosis is not a predictor of poor outcomes in primary total hip arthroplasty: a systematic literature review. International orthopaedics. 2011;35:465-73.
388. Conroy JL, Whitehouse SL, Graves SE, Pratt NL, Ryan P, Crawford RW. Risk factors for revision for early dislocation in total hip arthroplasty. The Journal of arthroplasty. 2008;23(6):867-72.
389. Donders ART, Van Der Heijden GJ, Stijnen T, Moons KG. A gentle introduction to imputation of missing values. Journal of clinical epidemiology. 2006;59(10):1087-91.
390. Eibich P, Dakin HA, Price AJ, Beard D, Arden NK, Gray AM. Associations between preoperative Oxford hip and knee scores and costs and quality of life of patients undergoing primary total joint replacement in the NHS England: an observational study. BMJ Open. 2018;8(4):e019477.
391. Teni FS, Burström K, Berg J, Leidl R, Rolfson O. Predictive ability of the American Society of Anaesthesiologists physical status classification system on health-related quality of life of patients after total hip replacement: comparisons across eight EQ-5D-3L value sets. BMC musculoskeletal disorders. 2020;21:1-13.
392. Oppe M, Devlin N, Black N. Comparison of the underlying constructs of the EQ-5D and Oxford Hip Score: implications for mapping. Value in Health. 2011;14(6):884-91.
393. Janssen M, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Quality of life research. 2013;22:1717-27.
394. Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Medical care. 2007:259-63.
395. Greene ME, Rader KA, Garellick G, Malchau H, Freiberg AA, Rolfson O. The EQ-5D-5L Improves on the EQ-5D-3L for Health-related Quality-of-life Assessment in Patients Undergoing Total Hip Arthroplasty. Clinical Orthopaedics and Related Research®. 2015;473(11):3383-90.
396. Lin DY, Cheok TS, Samson AJ, Kaambwa B, Brown B, Wilson C, et al. A longitudinal validation of the EQ-5D-5L and EQ-VAS stand-alone component utilising the Oxford Hip Score in the Australian hip arthroplasty population. Journal of Patient-Reported Outcomes. 2022;6(1):71.
397. Schatz C, Klein N, Marx A, Buschner P. Preoperative predictors of health-related quality of life changes (EQ-5D and EQ VAS) after total hip and knee replacement: a systematic review. BMC Musculoskelet Disord. 2022;23(1):58.

398. Huber M, Kurz C, Leidl R. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. BMC Med Inform Decis Mak. 2019;19(1):3.
399. Langenberger B, Schrednitzki D, Halder AM, Busse R, Pross CM. Predicting whether patients will achieve minimal clinically important differences following hip or knee arthroplasty. Bone & Joint Research. 2023;12(9):512-21.
400. Sullivan PW, Lawrence WF, Ghushchyan V. A national catalog of preference-based scores for chronic conditions in the United States. Medical care. 2005:736-49.
401. Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Loucks L, Khudairy AA, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. Quality of Life Research. 2015;24:1775-84.
402. Eneqvist T, Nemes S, Kärrholm J, Burström K, Rolfson O. How do EQ-5D-3L and EQ-5D-5L compare in a Swedish total hip replacement population? Acta Orthopaedica. 2020;91(3):272-8.
403. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-15.
404. Copay AG, Subach BR, Glassman SD, Polly Jr DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. The Spine Journal. 2007;7(5):541-6.
405. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Quality of life research. 2005;14:1523-32.
406. Cohen J. Statistical power analysis for the behavioral sciences: Academic press; 2013.
407. Kang S. Assessing responsiveness of the EQ-5D-3L, the Oxford Hip Score, and the Oxford Knee Score in the NHS patient-reported outcome measures. Journal of Orthopaedic Surgery and Research. 2021;16:1-12.
408. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively: a prospective cohort study of 1,335 patients. Acta orthopaedica. 2014;85(1):39-48.
409. Maillot C, Harman C, Al-Zibari M, Sarsam K, Rivière C. Moderate relationship between function and satisfaction of total hip arthroplasty patients: a cross sectional study. Hip International. 2022;32(1):25-31.
410. Abdel M, Watts C, Houdek M, Lewallen D, Berry D. Epidemiology of periprosthetic fracture of the femur in 32 644 primary total hip arthroplasties: a 40-year experience. Bone Joint J 98: 461–467. DOI 10.1302/0301-620X. 98B4. 37201; 2016.
411. Watts CD, Abdel MP, Lewallen DG, Berry DJ, Hanssen AD. Increased risk of periprosthetic femur fractures associated with a unique cementless stem design. Clinical Orthopaedics and Related Research®. 2015;473:2045-53.
412. Zhang Z, Zhuo Q, Chai W, Ni M, Li H, Chen J. Clinical characteristics and risk factors of periprosthetic femoral fractures associated with hip arthroplasty: a retrospective study. Medicine. 2016;95(35).
413. Chatziagorou G, Lindahl H, Garellick G, Kärrholm J. Incidence and demographics of 1751 surgically treated periprosthetic femoral fractures around a primary hip prosthesis. HiP international. 2019;29(3):282-8.
414. Scalici G, Boncinelli D, Zanna L, Buzzi R, Antonucci L, Di Maida F, et al. Periprosthetic femoral fractures in Total Hip Arthroplasty (THA): a comparison between osteosynthesis and revision in a retrospective cohort study. BMC Musculoskeletal Disorders. 2022;23(1):1-9.

415. Nugent M, Young SW, Frampton CM, Hooper GJ. The lifetime risk of revision following total hip arthroplasty: a new Zealand joint registry study. The bone & joint journal. 2021;103(3):479-85.
416. Lübbeke A, Roussos C, Barea C, Köhnlein W, Hoffmeyer P. Revision total hip arthroplasty in patients 80 years or older. J Arthroplasty. 2012;27(6):1041-6.
417. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. Statistics in medicine. 2013;32(16):2837-49.
418. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41-55.
419. Tsai M-C, Ng Y-Y, Chen W-M, Tsai S-W, Wu S-C. The effects of cement fixation on survival in elderly patients with hip hemiarthroplasty: a nationwide cohort study. BMC Musculoskeletal Disorders. 2019;20(1):1-8.
420. Lindberg-Larsen M, Petersen PB, Jørgensen CC, Overgaard S, Kehlet H, Hip LFCfF-t, et al. Postoperative 30-day complications after cemented/hybrid versus cementless total hip arthroplasty in osteoarthritis patients> 70 years: a multicenter study from the Lundbeck Foundation Centre for Fast-track Hip and Knee replacement database and the Danish Hip Arthroplasty Register. Acta orthopaedica. 2020;91(3):286-92.
421. Atik O, Çankaya D. To cement or not to cement, that is the question in elderly! Jt Dis Relat Surg. 2021;32(2):277-8.
422. Pedersen AB, Mailhac A, Garland A, Overgaard S, Furnes O, Lie SA, et al. Similar early mortality risk after cemented compared with cementless total hip arthroplasty for primary osteoarthritis: data from 188,606 surgeries in the Nordic Arthroplasty Register Association database. Acta orthopaedica. 2020;92(1):47-53.
423. Pedersen A, Baron J, Overgaard S, Johnsen S. Short-and long-term mortality following primary total hip replacement for osteoarthritis: a Danish nationwide epidemiological study. The Journal of Bone & Joint Surgery British Volume. 2011;93(2):172-7.
424. Moskal JT, Capps SG, Scanelli JA. Still no single gold standard for using cementless femoral stems routinely in total hip arthroplasty. Arthroplast Today. 2016;2(4):211-8.
425. Thalody HS, Post ZD, Bridges TN, Qadiri QS, Scaramella A, Ong AC, et al. Does Automated Impaction Improve Femoral Component Sizing and Alignment in Total Hip Arthroplasty? The Journal of Arthroplasty. 2023;38(10):2154-8.
426. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet. 2007;370(9597):1508-19.
427. Pennington M, Grieve R, Sekhon JS, Gregg P, Black N, van der Meulen JH. Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis. Bmj. 2013;346.
428. Galia CR, Diesel CV, Guimarães MR, Ribeiro TA. Total hip arthroplasty: a still evolving technique. Rev Bras Ortop. 2017;52(5):521-7.
429. Grassi L, Väänänen SP, Yavari SA, Jurvelin JS, Weinans H, Ristinmaa M, et al. Full-field strain measurement during mechanical testing of the human femur at physiologically relevant strain rates. J Biomech Eng. 2014;136(11).
430. Krull A, Morlock M, Bishop N. Maximizing the fixation strength of modular components by impaction without tissue damage. Bone & joint research. 2018;7(2):196-204.
431. Rehmer A, Bishop NE, Morlock MM. Influence of assembly procedure and material combination on the strength of the taper connection at the head–neck junction of modular hip endoprostheses. Clinical Biomechanics. 2012;27(1):77-83.
432. Sakai R, Takahashi A, Takahira N, Uchiyama K, Yamamoto T, Uchida K, et al. Hammering force during cementless total hip arthroplasty and risk of microfracture. Hip Int. 2011;21(3):330-5.

433. Cummins F, Reilly PO, Flannery O, Kelly D, Kenny P. Defining the impaction frequency and threshold force required for femoral impaction grafting in revision hip arthroplasty. A human cadaveric mechanical study. Acta Orthop. 2011;82(4):433-7.
434. Flannery OM, Britton JR, O'Reilly P, Mahony N, Prendergast PJ, Kenny PJ. The threshold force required for femoral impaction grafting in revision hip surgery. Acta Orthop. 2010;81(3):303-7.
435. Cristofolini L, Juszczyk M, Taddei F, Viceconti M. Strain distribution in the proximal human femoral metaphysis. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine. 2009;223(3):273-88.
436. Nicolella DP, Nicholls AE, Lankford J, Davy DT. Machine vision photogrammetry: a technique for measurement of microstructural strain in cortical bone. J Biomech. 2001;34(1):135-9.
437. Dickinson A, Taylor A, Ozturk H, Browne M. Experimental validation of a finite element model of the proximal femur using digital image correlation and a composite bone model. Journal of biomechanical engineering. 2011;133(1).
438. Dong B, Zeng F, Pan B. A Simple and Practical Single-Camera Stereo-Digital Image Correlation Using a Color Camera and X-Cube Prism. Sensors (Basel). 2019;19(21).
439. Pan B, Qian K, Xie H, Asundi A. TOPICAL REVIEW: Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. Measurement Science and Technology. 2009;20:062001.
440. Palanca M, Tozzi G, Cristofolini L. The use of digital image correlation in the biomechanical area: a review. International Biomechanics. 2016;3(1):1-21.
441. Rossman T, Uthamaraj S, Rezaei A, McEligot S, Giambini H, Jasiuk I, et al. A Method to Estimate Cadaveric Femur Cortical Strains During Fracture Testing Using Digital Image Correlation. JoVE. 2017(127):e54942.
442. Nwankwo CD, Parrish R, Leasure J, McGann WA. Prophylactic Cerclage With Braided Polyblend Suture During Femoral Broaching. Orthopedics. 2016;39(6):e1183-e7.
443. Ruspi ML, Palanca M, Faldini C, Cristofolini L. Full-field in vitro investigation of hard and soft tissue strain in the spine by means of Digital Image Correlation. Muscles Ligaments Tendons J. 2017;7(4):538-45.
444. Doyle R, van Arkel RJ, Jeffers JRT. Effect of impaction energy on dynamic bone strains, fixation strength, and seating of cementless acetabular cups. J Orthop Res. 2019;37(11):2367-75.
445. Reu P. Points on Paint. Experimental Techniques. 2015;39(4):1-2.
446. Aquarius R, Walschot L, Buma P, Schreurs BW, Verdonschot N. In vitro testing of femoral impaction grafting with porous titanium particles: a pilot study. Clin Orthop Relat Res. 2009;467(6):1538-45.
447. Doyle R, Arkel RJv, Muirhead-Allwood S, Jeffers JRT. Impaction technique influences implant stability in low-density bone model. Bone & Joint Research. 2020;9(7):386-93.
448. Doyle R, Boughton O, Plant D, Desoutter G, Cobb JP, Jeffers JRT. An in vitro model of impaction during hip arthroplasty. J Biomech. 2019;82:220-7.
449. Väänänen SP, Amin Yavari S, Weinans H, Zadpoor AA, Jurvelin JS, Isaksson H. Repeatability of digital image correlation for measurement of surface strains in composite long bones. J Biomech. 2013;46(11):1928-32.
450. Research V. Phantom Camera Hardware User Manual New Jersey: Vision Research; 2018 [Available from: https://phantomhighspeed-service.force.com/servlet/servlet.FileDownload?file=00P1N00000dSu7aUAC&oid=00D1N000002Br2EUAS.
451. Carriero A, Abela L, Pitsillides AA, Shefelbine SJ. Ex vivo determination of bone tissue strains for an in vivo mouse tibial loading model. J Biomech. 2014;47(10):2490-7.

452. Kim Y, Liu D, Lee H, Liu R, Sengupta D, Park S. Investigation of stress in MEMS sensor device due to hygroscopic and viscoelastic behavior of molding compound. IEEE Transactions on Components, Packaging and Manufacturing Technology. 2015;5(7):945-55.
453. Collis DK. Hip Replacement: Current Trends and Controversies. Journal of Bone and Joint Surgery, American Volume. 2003;85:587-8.
454. Silva P, Rosa RC, Shimano AC, de Paula FJ, Volpon JB, Defino HL. TAPPING PILOT HOLE: MECHANICAL ANALYSIS OF SHEEP VERTEBRA AND THE ARTIFICIAL BONE MODEL. Rev Bras Ortop. 2010;45(3):290-4.
455. Mold JW, Hamm RM, McCarthy LH. The law of diminishing returns in clinical medicine: how much risk reduction is enough? J Am Board Fam Med. 2010;23(3):371-5.
456. Deuel CR, Jamali AA, Stover SM, Hazelwood SJ. Alterations in femoral strain following hip resurfacing and total hip replacement. J Bone Joint Surg Br. 2009;91(1):124-30.
457. Hart NH, Nimphius S, Rantalainen T, Ireland A, Siafarikas A, Newton RU. Mechanical basis of bone strength: influence of bone material, bone structure and muscle action. J Musculoskelet Neuronal Interact. 2017;17(3):114-39.
458. Tang T, Cripton PA, Guy P, McKay HA, Wang R. Clinical hip fracture is accompanied by compression induced failure in the superior cortex of the femoral neck. Bone. 2018;108:121-31.
459. Mirza SB, Dunlop DG, Panesar SS, Naqvi SG, Gangoo S, Salih S. Basic science considerations in primary total hip replacement arthroplasty. Open Orthop J. 2010;4:169-80.
460. Elfar J, Menorca RM, Reed JD, Stanbury S. Composite bone models in orthopaedic surgery research and education. J Am Acad Orthop Surg. 2014;22(2):111-20.
461. Hetaimish BM. Sawbones laboratory in orthopedic surgical training. Saudi Med J. 2016;37(4):348-53.
462. Sun C, Liu H, Shang Y, Chen S, Yu Q. Scheimpflug camera-based stereo-digital image correlation for full-field 3D deformation measurement. Journal of Sensors. 2019;2019:1-11.
463. Lionello G, Sirieix C, Baleani M. An effective procedure to create a speckle pattern on biological soft tissue for digital image correlation measurements. Journal of the Mechanical Behavior of Biomedical Materials. 2014;39:1-8.
464. Singh A, Scholze M, Hammer N. On the influence of surface coating on tissue biomechanics – effects on rat bones under routine conditions with implications for image-based deformation detection. BMC Musculoskeletal Disorders. 2018;19(1):387.
465. Shen VC, Bumgardner C, Actis L, Park J, Moyer D, May-Nikstaitis K, et al. In Situ Deformation of First Tarsometatarsal Arthrodesis Implants with Digital Image Correlation: A Cadaveric Study. JOM. 2022;74(9):3357-66.
466. Hensley S, Christensen M, Small S, Archer D, Lakes E, Rogge R. Digital image correlation techniques for strain measurement in a variety of biomechanical test models. Acta Bioeng Biomech. 2017;19(3):187-95.
467. Huang L, Korhonen RK, Turunen MJ, Finnilä MAJ. Experimental mechanical strain measurement of tissues. PeerJ. 2019;7:e6545.
468. Bieger R, Ignatius A, Reichel H, Dürselen L. Biomechanics of a short stem: In vitro primary stability and stress shielding of a conservative cementless hip stem. J Orthop Res. 2013;31(8):1180-6.
469. Freitag T, Bieger R, Kiefer H, Dornacher D, Reichel H, Ignatius A, et al. Biomechanics of a calcar loading and a shortened tapered femoral stem: Comparative in-vitro testing of primary stability and strain distribution. J Exp Orthop. 2021;8(1):74.

470. Naghavi SA, Lin C, Sun C, Tamaddon M, Basiouny M, Garcia-Souto P, et al. Stress Shielding and Bone Resorption of Press-Fit Polyether-Ether-Ketone (PEEK) Hip Prosthesis: A Sawbone Model Study. Polymers (Basel). 2022;14(21).
471. Kim YH, Kim JS, Cho SH. Strain distribution in the proximal human femur. An in vitro comparison in the intact femur and after insertion of reference and experimental femoral stems. J Bone Joint Surg Br. 2001;83(2):295-301.
472. Cristofolini L, Juszczyk M, Taddei F, Viceconti M. Strain distribution in the proximal human femoral metaphysis. Proc Inst Mech Eng H. 2009;223(3):273-88.
473. Otani T, Whiteside LA, White SE. Strain distribution in the proximal femur with flexible composite and metallic femoral components under axial and torsional loads. J Biomed Mater Res. 1993;27(5):575-85.
474. Li J-j, Tian D-m, Yang L, Zhang J-y, Hu Y-c. Influence of a metaphyseal sleeve on the stress-strain state of a bone-tumor implant system in the distal femur: an experimental and finite element analysis. Journal of Orthopaedic Surgery and Research. 2020;15(1):589.
475. Naghavi SA, Tamaddon M, Garcia-Souto P, Moazen M, Taylor S, Hua J, et al. A novel hybrid design and modelling of a customised graded Ti-6Al-4V porous hip implant to reduce stress-shielding: An experimental and numerical analysis. Front Bioeng Biotechnol. 2023;11:1092361.
476. Bishop NE, Wright P, Preutenborbeck M. A parametric numerical analysis of femoral stem impaction. PLoS One. 2022;17(5):e0268561.
477. Bülow E. coder: An R package for code-based item classification and categorization. Journal of Open Source Software. 2020;5(56):2916.
478. Logishetty K, Gofton WT, Rudran B, Beaulé PE, Cobb JP. Fully Immersive Virtual Reality for Total Hip Arthroplasty: Objective Measurement of Skills and Transfer of Visuospatial Performance After a Competency-Based Simulation Curriculum. JBJS. 2020;102(6):e27.
479. Lewis S, Glen J, Dawoud D, Dias S, Cobb J, Griffin X, et al. Venous Thromboembolism Prophylaxis Strategies for People Undergoing Elective Total Hip Replacement: A Systematic Review and Network Meta-Analysis. Value in Health. 2019;22(8):953-69.
480. (NIHR) NIfHaCR. Thromboprophylaxis in Lower Limb Immobilisation (TiLLI): a study comprising two linked randomised controlled trials evaluating the effectiveness and cost effectiveness of different methods of pharmacological prophylaxis for patients with temporary lower limb immobilisation 2023 [Available from: https://fundingawards.nihr.ac.uk/award/NIHR154716.
481. VentureRadar. ADDITIVE INSTRUMENTS LIMITED [Available from: https://www.ventureradar.com/organisation/ADDITIVE%20INSTRUMENTS%20LIMITED/2133b7d5-d0e8-4981-9cd9-7bca829fcca1.