Statistical Approaches to the Early Assessment of Hip and Knee Replacement Prostheses

by

Khashayar Ghadirinejad

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy

College of Science and Engineering

June 2022

Table of Contents

Abstract

Joint replacement is a highly successful and frequent surgical intervention. It can improve function and reduce pain in patients with end-stage arthritis of the joints. However, there is a wide variation in the outcome of prostheses/devices used in primary total hip replacements (THRs) and primary total knee replacements (TKRs). Joint replacement registries have significant roles in assessing the comparative performance of devices. The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) has established a standardised multi-stage approach for identifying prostheses with a higher than anticipated revision rate, also referred to as 'outliers'. The AOANJRR standard compares the revision rate of prostheses to the average revision rate of all other prostheses that belong to the same broad device class—comparator. However, as changes are made in the design and performance of devices over time, the hip and knee comparator classes need to be re-evaluated. This study first aimed to explore how the rate of revision estimated in the comparator groups differs according to specific prosthesis design constructs. The cumulative percent revision (CPR) was calculated for 413,417 primary THR and 640,045 TKR undertaken for osteoarthritis from 1st January 2003 to 31st December 2019. The final hip comparator, which only includes satisfactory-performed prostheses of contemporary design and use, had a 10-year CPR of 4.30% (4.2, 4.41) which is lower than the current THR comparator used by the AOANJRR of 4.93% (4.84, 5.02). Using a comparator that only includes contemporary devices with modern bearings and excludes special devices used in more complex primary procedures has the potential to improve the early assessment of modern primary total hip prostheses sensitively. The AOANJRR standard detected 13 additional total conventional hip components utilizing the modified comparator. The final comparator group for conventional TKRs, which only includes the Cruciate Retaining and Posterior Stabilised designs, indicated a 10-year CPR of 5.2% (5.1, 5.3). Moreover, a comparator group of complex knee devices with 10.3% (8.6, 12.0) 10-year CPR was explored to reflect devices used only for specific purposes in primary TKR. The use of modified knee comparator groups led to identifying additional conventional knee prostheses but fewer complex knee designs as being at risk. The AOANJRR currently recommends the modern comparator groups for the early assessment of primary total hip and knee prostheses. Ideally, early identification of outliers uses a time-to-event outcome while reducing the confounding effects of other components in the device and patient characteristics. Machine learning (ML), which contains self-learning algorithms, is one approach to consider many variables simultaneously to reduce the impact of confounding. Another principal objective of this study was to compare the effectiveness of using either Random Survival Forest (RSF) or regularized/unregularized Cox regression to account for patient and associated device confounding factors to current standard techniques. The effectiveness of the ML approaches was assessed based on the ability to detect the outliers identified by the AOANJRR standardised approach, where the standard identified ten individual THR prostheses and five TKR prosthesis combinations. The ML approaches identified some but not all the outliers detected by the AOANJRR in the study cohort. Both the methods identified three of the same THR prostheses, and the RSF identified the other five of the detected THR components. In primary TKR, both feature selection techniques identified two of the same total knee prostheses, and Cox detected one additional prosthesis as at higher risk of revision. In addition, both the RSF and Cox techniques detected a number of additional device components that were not previously identified by the standard approach. The results showed ML might be able to offer a supplementary approach to enhance the early identification of outlier devices. RSF was a more comparable feature selection technique to the AOANJRR standard. Further studies are required to better understand the potential of ML to improve the early identification of outliers.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Full Name: Khashayar Ghadirinejad

Signed: Khashayar Ghadirinejad *Date: 22/06/2022*

Acknowledgments

I would like to express my most profound appreciation to my principal supervisor, Dr. Reza Hashemi who patiently helped, supervised and supported me throughout my Ph.D. This work could not have been possible without his guidance and consistent help. I greatly appreciate my co-supervisor, Prof. Mark Taylor, whose thorough knowledge of the design of prostheses used in hip and knee replacements improved the quality of this research. I thank my adjunct supervisor, Prof. Bogdan Solomon, who has always been ready to help me at different stages of this work and supported me significantly in linking with the AOANJRR to access the data.

I acknowledge the AOANJRR, the patients, surgeons and hospitals whose data made this research possible. I would like to express my gratitude to Professor Stephen Graves (the Director of the AOANJRR), who offered continual support for the study design and gave me a fantastic opportunity to work with the professional staff at AOANJRR which is a leader in developing new biostatistical methods. I also want to say many thanks to Prof. Richard de Steiger and Dr. Nicole Pratt, who kindly helped me to overcome many of the research challenges by sharing their extensive knowledge of joint arthroplasty and statistical analyses.

Finally, I extend my special thanks to my wonderful family. Words cannot express my gratitude for all of the sacrifices they have made. They were always my support when there was no-one to respond to my thoughts and doubts.

List of Figures

List of Tables

Introduction

Total joint replacement surgery is commonly performed on patients with severe joint osteoarthritis (OA). However, there are concerns about prostheses being at increased risk of failure. The early identification of these prostheses can be a challenge because of the many distinct components available, and it is a complicated process to estimate their performance in the presence of prosthesis-, patient- and surgeon-related confounders. In orthopedics, joint registries (JRs) collect and record data of joint replacements to observe the survival rate of prostheses. The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) dataset shows that most prostheses have comparable outcomes, although some have a higher-than-expected revision rate (called outliers). Machine Learning (ML)-based models are increasingly being used in the medical field to identify risk factors and possible outcomes. In orthopedics, ML methods can play a role in detecting components that are at an increased risk of failure and can be a useful tool for the initial screening of prosthesis components. First, the aim of this study was to improve the sensitivity of conventional analyses by identifying modified comparator groups to detect outliers as early as possible and with a high confidence level. The second aim was to evaluate the ability of feature selection techniques to identify outlier prostheses based on the historical data. It is anticipated that the findings of this research in regard to statistical techniques and the identification of outlier prostheses will be a significant step toward reducing the risk of using poor-performing prostheses implanted in large numbers of patients around the world.

Chapter 1: A comprehensive literature review of the previous research on total hip replacement (THR) and total knee replacement (TKR) is provided. This includes sections on the causation, outcome, reason for revision, and recent advances made in hip and knee replacements. We also review the studies on the potential risk factors and the role of joint registries in monitoring the performance of prostheses. The review reveals a need to examine novel statistical methodologies to improve the outlier identification process by JRs.

Chapter 2: This chapter reviews the use of supervised ML techniques in regression, classification and survival problems associated with the post-operative outcomes of THR and TKR. The different types of ML techniques such as Random Forest, SVM, Naive Bayes, and Deep Learning were reviewed, focusing on the data source, domains, limitations, and the quality of reported outcomes. The literature shows that ML adoption for post-operative THR and TKR is still in the emergent phase, offering opportunities for ML-based research studies in this area.

Chapter 3: The chapter includes an investigation of a number of different comparator options to provide a more relevant standard for comparing the performance of new hip devices. Subsequently, the current comparator was upgraded to adequately reflect contemporary practices and avoid an overestimation of the revision rate. The AOANJRR standard detected 13 additional device components utilizing the modified comparator. The registry currently recommends the modern comparator for early assessment of total hip prostheses.

Chapter 4: The use of ML methods (random survival Forest (RSF) and regularised/unregularized Cox regression) was evaluated for surveillance of total hip arthroplasty components. Their effectiveness was determined based on their ability to detect the same hip outliers identified by the AOANJRR standard using the comparator developed in chapter three. Both the RSF and Cox techniques detected a number of additional device components not found through the standardised approach, and RSF was a more comparable feature selection technique to the AOANJRR standard. Machine Learning might be able to offer a supplementary approach to improve the early identification of outlier devices.

Chapter 5: Given the higher revision risk of complex knee designs in primary total knee surgeries, this chapter was designed to identify more relevant comparator groups to better reflect conventional and complex surgical practices. Conventional designs include cruciate retaining (CR) and posterior stabilised (PS), and the knee designs used in more complex surgery include fully stabilised (FS) and hinged designs. The CR and PS groups were combined to produce the final conventional comparator. In addition, the FS and hinged designs were combined as a comparator group of complex devices. The use of modified comparator groups led to the identification of additional conventional prostheses but fewer complex designs as being at risk.

Chapter 6: The chapter assessed the ability of the algorithms developed in chapter four to identify total knee outliers among 160 unique prosthesis combinations. These outliers were detected using the modified comparator groups defined in the previous chapter. The two-step Cox yielded outcomes similar to those of the RSF but had significant advantages in terms of the number of prostheses identified, computational time, interpretation of variable importance, and reduction of confounding effects. Cox modelling is a more conventional method of selecting significant variables and documenting the confounding. Further studies are required to better understand the potential of ML to improve the early identification of outliers.

Chapter 7: This chapter includes further investigations of the outlier prostheses similar to those identified by all the statistical approaches in terms of clinically-known confounding factors. The impact of design- and patient-related variables was examined closely to determine the factors contributing to the poor performance of outlier prostheses. There were significant differences in the survival outcomes of the identified prostheses in terms of bearing surface and fixation method.

Chapters 8 discusses the main findings and limitations, and chapter 9 provides conclusion remarks and recommendations for future research.

Chapter 1. Literature Review on Hip and Knee Replacement

1.1 Overview

Total hip replacement (THR) is more likely than any other elective surgical procedure to improve a patient's quality of life [1]. Implant technology has continuously advanced since the pioneering work of Wiles [2], Charnley, and others in the mid-20th century [3]. Nowadays, over 95% of artificial hip joints last longer than 10 years, far longer than predicted by Charnley. Total knee replacement (TKR) surgery has been performed consistently for more than 40 years, and its popularity worldwide is increasing [4]. It has had proven success in reducing pain and improving the long-term knee function in people with arthritis. However, some patients are unhappy with the results of their hip and/or knee surgery; hence, this field has been the focus of research and development. On one hand, the medical device industry continues to develop new implants and supporting technology, although more rigorous evidence is still needed to justify their products. On the other hand, enhanced rehabilitation programmes are becoming more common, given their potential to improve patient outcomes.

1.2 Causation

In Australia, Osteoarthritis (OA) is the leading cause of medical issues that ultimately require total hip and/or knee replacements, accounting for 88.2% and 97.7% of primary total procedures in 2020, respectively [5]. Osteoarthritis is caused by a combination of biochemical and mechanical processes that are influenced by hereditary and environmental variables [6]. Age, gender, trauma, and joint morphology are all important patient-specific risk factors. Femoroacetabular impingement is becoming more widely recognised as a cause of hip pain [7]. For unknown reasons, the relationship between obesity and hip OA is substantially weaker than obesity and knee OA [8]. Moreover, there is no convincing evidence that there is a link between OA and diet. As the world′s population ages, the prevalence of OA is expected to rise. In Australia, the mean age of patients requiring primary THR for all diagnoses is 67.7 years and 68.5 years for the knee. Primary total hip and knee replacement for all diagnoses are more common in women than men [5].

1.3 Assessment of Outcome

Kaplan-Meier survival analysis with revision surgery as the endpoint is the most common method for evaluating hip and knee surgery outcomes. A revision occurs when one or more components of a prosthetic hip or knee are replaced. The revision treatment is recommended only when serious symptoms, such as pain or fracture, appear or are expected [5]. This is because a revision can result in major complications and poorer functional outcomes than first hip or knee surgery.

Joint replacement registries are valuable sources for measuring the rate of implant revision. Since the establishment of the first hip arthroplasty registry in Sweden, it has been successful in identifying devices with significant failure rates [9]. The International Society of Arthroplasty Registries (ISAR) now includes members from arround 25 nations, indicating that geographic coverage has gradually expanded. In Australia, since 1999, the registry has reported an overall 19-year implant survival of 89.4% (95% CI, 92.6–92.9) for 421,141 primary total conventional hip replacements (excluding resurfacing procedures) and 91% for 699,283 primary total knee for OA [10].

The collection of additional data enables comparisons to be made regarding the effects of patient, procedure, hospital, and surgeon variables. Currently, the outcomes of revision procedures performed by individual surgeons are not being documented. It should be highlighted that revision surgery as the sole measure of success has limits because patients can experience problems, discomfort, or poor function without undergoing a revision. Patient-reported outcome scores are used alongside with survival outcomes to better reflect pain, function, quality of life, and satisfaction after joint replacements. The Oxford Hip Score which measures pain and functional status, and the EuroQol five-domain score that analyses the quality of life, are two patient-reported outcomes that are now regularly recorded for hip surgery [11, 12].

Although only a few of these approaches have been assessed for reliability, validity, and responsiveness, there has been a significant increase in the number of knee instruments and rating scales developed to measure outcomes from the patient's perspective. The Western Ontario and McMaster Universities Osteoarthritis (WOMAC), the Knee Injury and Osteoarthritis Outcome Score, and the Oxford Knee Score (OKS) were all frequently utilised in a recent systematic review [13]. However, because age and comorbidities influence patient-reported outcomes, it is impossible to have a universal threshold as a discriminator of success [14].

Joint replacement places a significant cost load on healthcare systems. For example, the annual cost of hip replacement in the United States exceeds $15 billion [15]. Hip replacement costs between $1,500 and $10,402 per quality-adjusted life year (QALY) gained [16, 17]. This figure is significantly lower than the £20000–30000 per QALY benchmark set by the National Institute for Health and Care Excellence (NICE) to guide cost-effectiveness assessments of novel technologies [18].

Total joint replacement surgery for the hips and knees is the most common inpatient operation for Medicare recipients, and the recovery time can be lengthy. However, data suggests that in the long run, arthroplasty saves money on healthcare. The healthcare costs for a patient with a hip or knee replacement are lower than those for a patient who does not have either prosthesis [19]. For patients with a reasonable life expectancy, a hip and/or knee replacement can be a cost-effective procedure.

1.4 Reasons for Revision

As indicated in Table 1.1, aseptic loosening is the most common reason for revision, accounting for 24.2% of all revisions of primary total conventional hip surgeries reported by the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR); this is followed by fracture (21.1%), prosthesis dislocation (20.3%), and infection (18.6%) [5].Wear of the bearing surfaces, which produces particle debris inside the enough joint space, is the most common cause of aseptic loosening. In registry records, the incidence of fretting and corrosion is definitely understated. Pain [20], infection [21, 22], and aseptic loosening of implants [23-27] can all be caused by the products of corrosion and particles of debris. Dislocations affect 0.2–10% of patients after hip replacement, with 77% of them occurring within the first year [28].

Table 1. 1. Primary total conventional hip for OA by reason for revision [5].

Reason for revision	Number	Percent
Loosening	3958	24.2
Fracture	3447	21.1
Prosthesis Dislocation	3329	20.3
Infection	3038	18.6
Lysis	357	2.2
Pain	322	2.0
Leg Length Discrepancy	257	1.6
Malposition	234	1.4
Instability	221	1.4
Implant Breakage Stem	192	1.2
Implant Breakage Acetabular Insert	151	0.9
Wear Acetabular Insert	142	0.9
Metal Related Pathology	141	0.9
Incorrect Sizing	106	0.6
Implant Breakage Acetabular	100	0.6
Implant Breakage Head	47	0.3
Other	327	2.0
Total	**16,369**	**100.0**

The dislocation rate is influenced by patients' age, muscle tone, failure to avoid particular movements, surgical approach, and the size and positioning of components [28]. Periprosthetic joint infection is a serious arthroplasty complication that can lead to pain, loss of function, systemic illness, and even death. Within two years of surgery, the frequency of such infection is 1–2% [29]. Biofilms are formed by microbes on implant surfaces, which reduce antibiotic penetration [30]. The most common reasons for revision vary depending on patient characteristics. Loosening is the most prevalent reason necessitating revision procedures for both males and females under 75 years of age, whereas fracture is the most common reason for revision in patients beyond 75 years of age [5].

Similar to the hip, aseptic loosening of the prosthesis is the most prevalent reason for a primary knee replacement to be revised [31-34], accounting for 24.7% of all revisions in Australia shown in Table 1.2 [5]. Implant wear is the most common cause of loosening. Because the rate of wear is a function of both time and activity, it is most concerning in young and energetic patients [35]. Infection is the second most common reason for primary knee revisions [34, 36], accounting for 23.7% of all primary knee revisions [5]. Although this dangerous outcome is frequently detected soon after surgery, it can occur at any time following the surgery [37, 38].

Table 1. 2. Primary total knee for OA by reason for revision [5].

Reason for revision	Number	Percent
Loosening	6805	24.7
Infection	6539	23.7
Patellofemoral Pain	2519	9.1
Instability	2345	8.5
Pain	2250	8.2
Patella Erosion	1645	6.0
Arthrofibrosis	990	3.6
Fracture	860	3.1
Malalignment	592	2.1
Lysis	541	2.0
Wear Tibial Insert	521	1.9
Metal Related Pathology	354	1.3
Incorrect Sizing	295	1.1
Other	1324	4.8
Total	**27,580**	**100.0**

Other reasons for revisions are pain following surgery, particularly patellofemoral pain, and instability which, along with loosening and infection, account for the majority of the revisions [5, 33, 34]. Although all of these reasons for revision may be related to the design or manufacture of implants, there are other factors that increase the likelihood of revision. Preoperative diagnosis, patient characteristics, the surgical procedure, the surgeon's experience and expertise, the facilities in the operating theatre, and postoperative care are other considerations [39].

According to a recent review of the New Zealand registry, early revision due to infection increased and similar patterns have been documented in Sweden and Australia. [5, 34, 40]. The rise in the number of patients with periprosthetic joint infection in the hip and knee has been attributed in part to an increase in the number of patients with diabetes or obesity or patients in younger age categories [41]. There is a need to improve the diagnosis of periprosthetic joint infection, and more research is also needed for better management [42]. The gathering of more relevant outcome data, such as microbiological profiles, antimicrobial medication, and the general health status of patients, is critical to this improvement, together with infection-specific outcomes from registry survival findings [43].

1.5 Advances in Hip Practice

In England, adults aged 65 to 74 spend an average of 6.5 hours a week engaged in physical activities [44]. There is a high demand for new hip implants, and the development of devices that can better withstand wear at the bearing interface, the femoral head and acetabular cup articulate, continues to be a major problem. In vivo, the ideal bearing surface is chemically inert, wears slowly, produces non-immunogenic wear debris, and is robust enough to withstand fracture. Implants with metal-on-polyethylene bearings, as well as ceramic-on-polyethylene and ceramic-on-ceramic bearings, were used in the majority of surgeries in Australia, shown in Table 1.3 [5]. Because the softer polythene generated wear-related debris, early models of metal-on-polyethylene bearings had a high failure rate at long follow-up times [5]. Current highly cross-linked polythene (XLPE) is more durable than the earlier materials, and the registry indicated no difference in mid-time revision rates for modern metal-on-polyethylene, ceramic-on-polyethylene, and ceramic-on-ceramic bearings [5, 45].

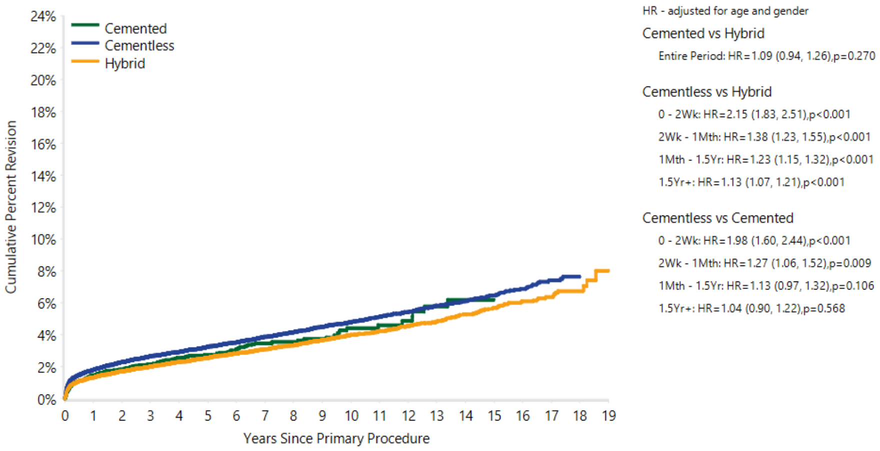
Table 1. 3. Percent revision of primary total conventional hip for OA by bearing surface [5].

Bearing Surface	N Revised	N Total	1 Yr	3 Yrs	5 Yrs	10 Yrs	15 Yrs	19 Yrs
Ceramic/Ceramic	3876	94733	1.5 (1.4, 1.6)	2.4 (2.3, 2.5)	3.1 (3.0, 3.2)	5.0 (4.8, 5.1)	7.1 (6.8, 7.4)	8.4 (7.9, 8.8)
Ceramic/Non XLPE	582	7986	1.9 (1.6, 2.3)	3.2 (2.8, 3.6)	3.8 (3.4, 4.3)	7.2 (6.5, 7.9)	11.8 (10.8, 12.9)	15.3 (13.9, 16.7)
Ceramic/XLPE	2484	91245	1.7 (1.6, 1.8)	2.5 (2.4, 2.6)	3.1 (2.9, 3.2)	4.3 (4.1, 4.6)	5.9 (5.4, 6.4)	7.5 (6.4, 8.8)
Ceramic/Metal	26	299	1.7 (0.7, 4.0)	3.7 (2.1, 6.6)	4.4 (2.6, 7.4)	8.4 (5.7, 12.3)		
Metal/Metal>32mm	3415	14422	1.7 (1.5, 1.9)	5.7 (5.3, 6.1)	11.8 (11.2, 12.3)	22.5 (21.8, 23.2)	28.5 (27.5, 29.5)	32.2 (29.1, 35.5)
Metal/Metal≤32mm	411	5146	1.6 (1.3,2.0)	3.3 (2.9, 3.8)	4.4 (3.9, 5.0)	6.7 (6.0, 7.4)	9.1 (8.3, 10.1)	10.1 (9.1, 11.3)
Metal/Non XLPE	2821	35266	1.4 (1.3, 1.6)	2.5 (2.3, 2.7)	3.5 (3.3, 3.7)	6.5 (6.2, 6.7)	11.0 (10.6, 11.4)	13.6 (13.0, 14.2)
Metal/XLPE	5792	165771	1.6 (1.6, 1.7)	2.4 (2.3, 2.5)	3.0 (2.9, 3.1)	4.6 (4.5, 4.7)	6.3 (6.1, 6.6)	7.3 (6.9, 7.8)
Cermicised Metal/Non XLPE	50	297	1.7 (0.7, 4.0)	3.8 (2.1, 6.7)	4.1 (2.4, 7.2)	12.5 (8.9, 17.3)	20.7 (15.7, 27.1)	
Cermicised Metal/XLPE	724	25323	1.8 (1.6, 2.0)	2.3 (2.1, 2.5)	2.6 (2.4, 2.9)	3.8 (3.5, 4.1)	5.5 (4.8, 6.3)	
Total	**20181**	**440487**						

Note. Excludes 213 procedures with unknown bearing surface, 1 procedure with ceramicised metal/ceramic bearing surface, 8 procedures with metal/ceramic bearing surface.

With only minor changes evident in the rate of revisions, various criteria may help surgeons determine which bearing to use. Although modern ceramic-on-ceramic bearings are more expensive than others and sometimes produce a squeaking sound, they do not have a higher risk of implant fracture compared to the previous, more brittle version [46]. Because metal-on-metal prostheses have less linear wear on the surface of the bearing compared with metal-on-polyethylene prostheses, they became popular 20 years ago. When registry data revealed the higher risk of metal-on-metal implants, this implantation peaked in 2008, accounting for 21% of all prostheses [5]. However, we know that cementless metal-on-metal THRs have a revision rate of more than 18% at 10 years [47].

The ideal method of fixation in THR is still a subject of debate (Figure 1.1). Cemented fixation has superior long-term performance and, it has a lower overall rate of revision than cementless fixation after 14 years. Cemented fixation continues to show excellent long-term revision rates and achieves a lower overall revision rate in longer times than cementless fixation [5, 47, 48]. It has also been found that higher failure rates of implants with cementless fixation indicate early fixation failure. Beyond the first decade of implantation, however, cementless fixation may have lower revision rates than cemented fixation [49]; This may result in a decreased revision rate in patients under the age of 65 [50].



Number at Risk	0 Yr	1 Yr	3 Yrs	5 Yrs	10 Yrs	15 Yrs	19 Yrs
Cemented	6751	5992	4655	3286	602	111	1
Cementless	220473	194040	147887	106008	30362	4908	28
Hybrid	117756	103480	79460	57955	20700	3979	50

Note: Includes mixed ceramic/mixed ceramic and cross-linked polyethylene (XLPE) bearing surfaces

Figure 1. 1. Percent revision of primary total conventional hip replacement for OA by fixation [5].

In Australia, the United Kingdom, and the United States, cementless fixing is the most popular. Acetabular aseptic loosening has been recognised as a primary cause of cemented implant failure in patients under the age of 60 [51]. Hybrid fixation, which combines cemented femoral and cementless acetabular components, was developed as an alternative and has outperformed the other fixations over a 15-year follow-up period [47]. Short cementless femoral stems are also gaining popularity. These designs maintain proximal bone stock and enable more physiological loading, which means less stress shielding, thigh pain, and invasive revision surgery.

In the last decade, there has been a tendency to increase the diameters of the femoral head, which reduces the risk of dislocation following hip replacement because a larger head enables a greater range of motion before impingement than does a smaller head [52]. Previous concerns about the use of large diameter heads stemmed from evidence that these implants cause increased polythene volumetric wear; however, with modern generations of highly cross-linked polythene, larger articulations do not appear to cause increased wear when compared to smaller articulations [53].

In the United Kingdom, posterior and lateral surgical approaches accounted for 95% of hip replacements, and both had excellent outcomes [47]. The posterior method has become more popular than the lateral approach in recent years. This rise may be due to mounting evidence that the posterior technique is linked to better patient-reported outcomes with no greater risk of dislocation [54, 55]. Interest in minimally invasive surgical methods has grown as a result of a desire to perform hip replacements with less soft tissue injury.

The direct anterior approach is one such technique. Despite early reports of better results, comprehensive studies indicated no significant difference in overall complication rate, dislocation rate, gait, or patient function after six weeks compared to other routine approaches. As yet there is no clear evidence showing how it affects fracture rate and the length of stay in the hospital [56, 57]. In New Zealand, the UK, and Sweden, the method is now used in less than 5% of procedures.

The percutaneously-assisted, super capsular, and direct superior approaches utilise a modified posterior incision and provide access to the joint without disturbing the external rotator muscles. Low rates of complications and dislocation have been documented in case studies [58]. All minimally invasive surgical procedures, however, require long-term monitoring. Impingement, increased surface wear of bearings, dislocation, and the need for revision can all result from the misalignment of acetabular and femoral components [59].

Computer-assisted surgical methods for hip replacement have been created in order to improve the accuracy and dependability of implant placement. From passive computer navigation to patient-specific instrumentation to active robotic-assisted surgery, there is a wide range of options. Computer navigation assists surgeons intraoperatively by using anatomic data from preoperative CT imaging, intraoperative fluoroscopic pictures, or imageless intraoperative registration of bone landmarks. In a meta-analysis of 473 patient data, researchers discovered that computer navigation improves the precision of the acetabular component positioning when compared non-navigated implantations [60]. However, as yet there is no evidence of better clinical outcomes.

In the case of knee replacement, research conducted in Australia suggests that computer guidance reduces the number of revisions in patients under the age of 65

[61]. The reasons for this disparity are unknown. Overall, robotic-assisted orthopaedic surgery involves procedures that are quite distinct from those used for soft-tissue surgery. The computer-assisted surgery system analyses an extensive amount of imaging data and searches for any deviations from a surgical plan. Some systems provide haptic feedback to surgeons to avoid the re-sectioning of bone beyond certain limits, while others automatically stop bone milling. Although the accuracy of acetabular positioning has improved, the impact on clinical outcomes has yet to be determined [62].

A single-center cohort study found that dislocation rates are lower following robotic-assisted hip replacement than a routine non-assisted hip replacement [63]. Three-dimensional templates produced from preoperative photographs are used to create a patient-specific apparatus. The procedures are designed to improve acetabular placement without the significant amount of time required for robotic surgery [64]. To determine the advantage of modified computer-assisted surgery systems, high-quality longitudinal studies are required.

1.6 Advances in Knee Practice

The outcomes of modern knee replacements are quite acceptable and are continuing to improve. The Swedish knee registry was able to track this development throughout its long history (Figure 1.2) [65]. With the current surgical procedures used for knee replacement, few patients require revision, especially in cases of young people. However, the patients who are not totally satisfied with the TKR outcome outnumber those who require revision. For most knee replacements, metal on polyethylene is still being used and, as a result, polyethylene wear remains a major reason for failure [34, 47, 66]. Highly cross-linked polythene, also known as second-generation polythene, was introduced around 20 years ago and successfully reduced polythene wear, thereby decreasing the rate of aseptic loosening and revision [67]. Recently, vitamin-E-infused highly cross-linked polythene, also known as a third generation development, has been used although its efficacy needs further investigation [68].

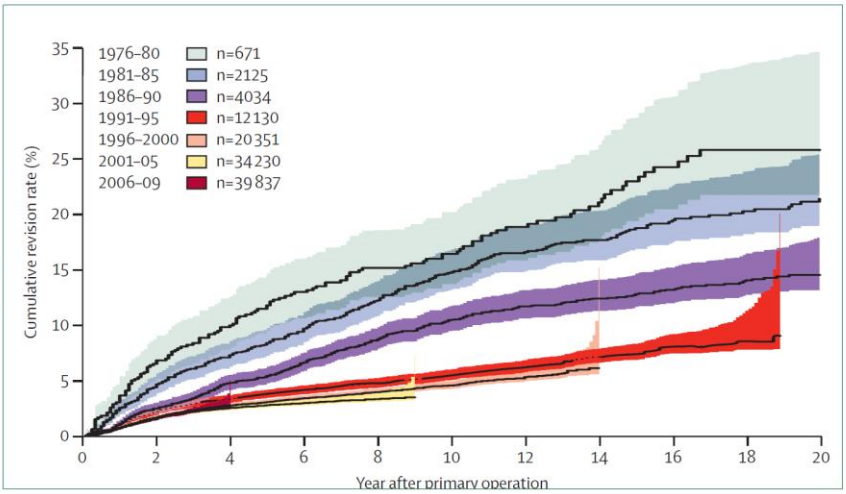


Figure 1. 2. Cumulative revision rate after TKR [65].

For more than 30 years, the standard method to TKR surgery has been to strive for mechanical alignment, which means that the hip, the center of the reconstructed knee, and the ankle are all in line [69]. Recently, the kinematic approach has been planned as an alternate implantation strategy to mimic the alignment of the pre-disease joint surface. This procedure is thought to improve ligament balance and knee kinematics without requiring ligament releases [70]. The global experience with kinematic alignment in TKR is limited [71], although a recent literature review found that kinematically-aligned surgery resulted in a better outcome than did mechanical alignment [72]. The pattern of OA may influence the benefit of different alignment methods for each individual patient. Mechanical alignment is still the most commonly used implantation method, and more research is needed on the safety of kinematic alignment before considering it for more widespread use.

The majority of patients receive total knee implants but about 8% of cases currently receive partial (unicompartmental) knee replacement [5]. Partial knee replacement has several advantages over total knee replacement: improved functional outcome, shorter postoperative length of stay, decreased mortality rate, and greater cost-effectiveness [73]. Recent evidence has reinforced these findings in randomised controlled studies, and similar results of current studies are being anticipated [74, 75]. The fundamental argument against widespread acceptance of partial knee as an alternative to TKR is that almost all the national registries report a greater revision rate [47, 66]. There is some evidence to suggest that the greater revision rate of partial knee replacements is related to the number of these surgeries performed by specific surgeons and private units [76]. In addition, recent evidence derived from registry data analysis indicates that the introduction of cementless fixation may decrease the revision rate of partial knee arthroplasty [40].

In knee replacement surgery, computer navigation and patient-specific instrumentation have been introduced to help reach a more accurate alignment [77]. Possible improvements are expected to produce better results and improve intraoperative efficiency and cost-effectiveness. Although the literature contains inconsistent reports, there is no significant clinical difference in terms of implant component placement, lower extremity alignment, or patient outcomes [64, 78]. The main advantage of patient-specific instrumentation or computer navigation lies in its potential to help less experienced surgeons achieve greater precision and accuracy. Some evidence suggests that computer-assisted surgery may reduce the revision rate for young patients undergoing total knee arthroplasty [61]. To date, there has been only a limited application of robotic technology in knee replacement surgery, and high-quality comparative studies are still needed to show that its potential efficacy is better than that of conventional techniques [78].

1.7 Joint Registries

Joint registries (JRs) are now one of the greatest and most essential sources of comparative data for hip and knee replacement surgery [79-81]. They can give us useful information about implants and show how patient variables, surgical procedures, and surgeon experience affect outcomes [5, 34, 47]. These JRs provide information on potential issues regarding implants and surgical procedures. Statistical studies of these registries are being conducted to discover problems with implants and surgical procedures. The Swedish hip and knee replacement registries were the leaders in this field, and their main goal was to collect data related to short-term problems. The registries quickly demonstrated their potential as a safe and reliable source, focusing on reporting outcomes related to the effectiveness of surgeries and prostheses.

The monitoring of revision rates is an acceptable way of determining the factor(s) affecting the survival outcome. This is especially practical in the field of joint replacement, where the goal has always been to improve the longevity of medical implants and reduce the incidence of failure. Significant international relationships between various registries and research groups should be established as one means of achieving this goal. Furthermore, the nationwide statistics published by registries can provide surgeons and patients with valuable information about what can be expected after surgery. The statistical analysis of registry data has become a key approach for evaluating the outcome of joint replacements associated with advancements in data collection [34, 40, 66]. In recent times, big data has been used in healthcare applications to develop prediction models and clinical decision support systems [82-86]. Hence, the mining of big data could yield valuable insights into the factors that contribute to the success or otherwise of hip and/or knee replacement surgery.

Overall, although most primary total knee prosthetic implants are well-established cruciate-retaining or posterior-stabilised devices with a long track record, new implant modifications or novel designs are being introduced regularly [5, 34, 47, 66]. Moreover, THR is mostly limited to the surgeries performed using a choice of modern bearings: XLPE with metal, ceramic or ceramicised metal heads, and ceramic on ceramic bearings with mixed ceramic [5]. Here, the importance of registries is underscored by introducing the Beyond Compliance in the UK, which is working narrowly with the Orthopaedic Data Evaluation Panel and the UK national joint replacement registry.

1.8 Potential Risk Factors in Hip and Knee Surgery

The most important patient-related factor influencing the outcome of hip and knee replacement is age. The majority of patients who require joint replacement surgery are elderly; the average age at the time of surgery is close to 70 years [5, 34, 47]. Because the incidence of OA increases with age [87], this finding is not unusual. Patients with OA and who are under 65 years old had 2.5 times more risk of knee revision than those who were 65 years or older [5, 34, 88]. When revision rates for various age groups are examined, it becomes clear that there is an inverse relationship between risk of revision and age, with younger patients having a higher risk of requiring revision surgery [5]. Other patient characteristics, such as preoperative diagnosis and gender, have also been shown to affect joint replacement surgery outcomes. In both hip and knee surgeries, men have a somewhat greater rate of revision than women [5, 34], particularly for infection ($p<0.001$) [5].

The AOANJRR is collecting data on other patient characteristics that could affect the outcome of hip and knee procedures, such as the American Society of Anaesthesiologists (ASA) score and Body Mass Index (BMI) [5]. A number of implant-related variables were described in the literature that may affect hip replacement failure rates, including head size, bearing material combination, and fixation method [5, 47, 66]. In regard to primary total knee risk factors, other variables can contribute to the outcome of surgeries: bearing mobility, stability, utilisation of patella component, and method of fixation [5, 47, 66].

The type and class of implants also have an impact on the result. The most prevalent type of primary partial hip is unipolar modular, which is defined by the class of prosthesis utilised [5]. This accounts for 45.6% of all partial hip surgeries, with two categories of unipolar monoblock (28.5%) and bipolar (25.9%). The major revision diagnostic for the three main classifications of primary partial hip replacement is a fractured neck of the femur, which is likely to have a greater rate of revision than primary total hip surgery [5]. All partial knee replacements (unicompartmental, patella and trochlear, and bicompartmental) have a greater revision rate than do the total knee [34, 40, 66]. Partial knee replacement is preferred over TKR for a variety of reasons, including a more natural feeling in the knee, less invasive surgery with a lower risk of infection [89, 90].

1.9 Prostheses with Higher Than Anticipated Rate of Revision

The majority of registries have recorded implants that have a much higher risk of revision than others within the same broad class. When considerable disparities in the revision rate for specific implants are observed, it is usually due to a problem with the design or the materials. When an implant has a higher risk of revision, its use decreases as surgeons subsequently choose other options. The implant may then be removed gradually from the market by reducing the usage.

Orthopaedic manufacturers and clinicians are constantly introducing new designs with a view to improving survival outcomes. The registries, which are the ultimate quality-assurance monitoring system, should track the results of novel implants and technologies. Based on the evidence [81, 91-93], many of these new-design implants are launched with little or no clinical evidence to support the manufacturers' claims.

The complex interaction of factors is becoming more apparent over the time. Understanding how these factors interact and how to appropriately weigh other variables will help surgeons to improve patient outcomes.

To detect an outlier prosthesis, the registries' procedures must be transparent and accountable. The various arthroplasty registries differ in terms of the methods used for the identification. The Swedish Hip Arthroplasty Register offers survival curves for prosthetic devices but does not provide any specific comparisons [34, 94]. The joint registry of New Zealand releases tables on prosthetic results but does not identify outlier prostheses [40]. The National Joint Registry for England, Wales, Northern Ireland and the Isle of Man (NJR) has formed an outlier subcommittee to explore strategy and methodology for analysing data on each implant specifically that has been identified as requiring further examination [47].

On September 1, 1999, the AOANJRR began a phased installation of data collection and had been registered complete national data since 2002. This registry has created a standardised three-stage process shortly for identifying prostheses with a higher-than-expected revision rate [66]. Stage 1 has been in place since the registry's inception, stage 2 was established in 2003, and stage 3 in 2007. The method by which the AOANJRR identifies prostheses with higher-than-expected revision rates has both advantages [94] and disadvantages [95].

Stage 1 involves a good screening test of prostheses; however, it does not take into account variations in revision rate over time. Because of this constraint, it is challenging to recognize a variance if the higher risk of revision happens later in the follow-up period [96]. The addition of stage 2 enabled more in-depth research of a variety of variables, both device- and non-device-related. Stage 3 has also proven to be beneficial because it expands the AOANJRR's clinical perspective. Because of the vast number of prostheses data submitted to the registry, it is impossible for the surgeons to be familiar with all of them. Surgeons involved in stage 3 have a sound knowledge of several of the devices and can add more detail to the registry's findings. This increases the Annual Report's transparency and accountability by assuring peer assessment of the data from a leading arthroplasty organisation.

1.10 Summary

Hip replacement surgery is still one of the most effective and reliable surgical procedures. Millions of patients with significant hip pain and functional limitations have subsequently been able to recover a much-improved quality of life. Implant material and design, surgical technique, and perioperative treatment have all improved. Most patients can expect their prostheses to last more than 20 years without causing issues. There are continued challenges that include improving implant function for active young patients, guaranteeing the safe introduction of novel implants, and developing techniques for the early detection of OA and control of its progression. The most significant advancements in modern total hip producers have been made by increasing the usage of only those devices that contain modern bearings.

Knee replacement surgery is a well-established procedure with a high rate of successful treatment outcomes and long-term implant survival. However, a percentage of patients continue to experience poor outcomes, and tackling this problem is a key challenge for improving treatment, especially in regard to the growing number of younger patients undergoing surgery. Because incremental changes in implant design have not significantly improved patient outcomes, more research is needed to optimise the performance of surgery based on unique patient characteristics.

Joint registries help progress the understanding of patient-reported knee replacement outcomes, but it still needs to be refined. National registries are helping us learn more about joint arthroplasty, and new analysis approaches must be used to derive the most information from the collected data. As with many medical fields, new technology is rapidly evolving in joint replacement, which may assist clinicians in the future to choose devices based on unique safety criteria for each specific patient. Innovative assessment approaches, including more randomised controlled trials and adaptable designs, are required prior to introducing new devices.

For many years, JRs have assessed outcomes after hip and knee replacement surgery. The established methodology estimates the time before a revision operation is required to identify poor-performing implants. Nevertheless, novel statistical methodologies such as machine learning algorithms can be suggested for future works using big data from JRs. Recent outcomes according to patient-reported data have also enabled the community to better evaluate functional results. A combination of revision and patient-reported measures as the endpoints could be a more practical indicator of the success or otherwise of an implant. The value of the registry data could be enhanced by conducting registry-based randomised controlled trials and improving the quality of research studies. Registered evidence can also directly contribute to healthcare delivery, as seen in Australia, where the AOANJRR reports are utilised routinely in individual consulting evaluations and hospital-level feedback.

Chapter 2. Literature Review on Supervised Machine Learning in Hip and Knee Replacements

2.1 Overview

Prediction models are being increasingly used in the medical field to identify risk factors and possible outcomes. Some of these are presently being used to develop guidelines for improving clinical practice. The application of Machine Learning (ML), comprising a powerful set of computational tools for analysing data, has been clearly expanding in the role of predictive modelling. This chapter reviews the latest developments of supervised ML techniques that have been used to analyse data related to post-operative total hip and knee replacements. The aim was to review the most recent findings of relevant published studies by outlining the methodologies employed (most-widely used supervised ML techniques), data sources, domains, limitations of predictive analytics and the quality of predictions.

Machine Learning (ML) provides systems the ability to produce mathematical models by learning patterns from empirical data. In medical research, ML is mostly used to extract information regarding diagnosis and treatment patterns. Examples include data-driven predictions of drug effects and interactions, the detection of comorbidity groups in autism spectrum disorders [97], and the identification of type 2 diabetes subgroups [98].

The most widely used ML approach in medical sciences is supervised learning. This technique estimates the mapping function for new input data in order to predict categorised, real values, or time-to-event outputs. Examples of supervised ML algorithms in orthopaedics include Linear Regression and similar techniques, Decision Trees (DTs), Random Forests (RFs), Neural Networks (NNs), Naive Bayes, Support Vector Machines (SVMs) and Nearest Neighbors [99].

ML technology is relatively new to the field of orthopaedic surgery. Recent applications of ML technology include the development of image-based diagnoses [100-102] and the improvement of value-based care [103-106]. Gait analysis algorithms may be used to notice early warning indications of revision arthroplasty, such as undiagnosed infection or instability [107]. Kotti et al. [102] used RF-based modelling to detect osteoarthritis (OA) through gait analysis, reporting a mean accuracy of 72% in 47 patients with this disease. In the area of value-based payment models, Navarro et al. [104] evaluated applying a Naïve Bayesian model to assess patient-level factors and forecast value metrics prior to the total knee arthroplasty episode of care. Similarly, Ramkumar et al. [103] explored the use of a Naïve Bayesian classifier in primary total hip arthroplasty, and found an excellent predictive capacity with respect to costs and hospital length of stay (LOS).

ML is increasingly used in the medical sciences because it offers alternative approaches to address the probability of confounding, particularly in high-dimensional datasets. For instance, Kaplan-Meier is a common statistical method that uses lifetime data to estimate the survival function of primary Total Hip Replacement (THR) and Total Knee Replacement (TKR) implants [79, 108-115]. However, an ideal method uses a time-to-event endpoint though reducing the confounding effect of other variables.

There are a number of important questions that need to be addressed given the recent advances and increasing use of supervised learning methods in various medical areas, including orthopaedics. These comprise: What are the main justifications for using supervised ML methods and their effectiveness in assisting with the THR and TKR procedures? Are the ML results affected by data volume and data quality?

The aim of this chapter was to address these questions by reviewing the use of supervised ML techniques in regression, classification and survival problems associated with the post-operative outcomes of THR and TKR. The different types of ML techniques (including RF, SVM, Naive Bayes, and Deep Learning) were reviewed, focusing on the data source, domains, limitations, and quality of outcomes reported in the literature.

2.2 Method (Literature search and selection criteria)

For this review, the English-based literature was searched online, including PubMed search engine and Scopus Elsevier databases using various key terms: supervised learning, machine learning, hip replacement, knee replacement, predictive, and data. A comprehensive search was conducted across these databases for the period of each database inception to the end of 2019. Only articles and their corresponding references reporting studies that utilised supervised ML techniques were reviewed for inclusion. Non-peer-reviewed studies, non-English language studies, unpublished manuscripts were excluded. The studies using unsupervised (or semi-supervised) ML learning approaches, or employing learning methods to train unreal data, or with a focus on pre-operative outcomes of THR and TKR were not considered. Titles and abstracts of the remaining articles were then carefully screened.

2.3 Random Forest (RF)

The RF is a tree-based ensemble learning method widely used to predict an outcome or rank and select the most significant variables. Cafri et al. [116] defined the time to first revision in elective primary THR as the outcome in order to compare two ML techniques (elastic-net VS. random survival forest) with the principal aim of assessing their performance in identifying recalled components. The concept of training an ML model to identify significant features differs from predicting the survival probability of components. The authors used 348 unique components as indicator variables in addition to patient covariates, which were all categorised and treated as potential confounders to detect the components based on the statistical significance. Two of the six recalled components (ASR shell/head and Rejuvenate) with $P<0.001$ and minimal depth rank of 1 and 2 in the RF model, were identified in both approaches. However, one more component (Durom shell/Metasul femoral head) was also picked by the regularized Cox model, even while maintaining the false discovery rate at .05. The results suggested that the ML methods can be effective for detection, although the Cox technique with a more traditional way to address confounding performed more effectively [116].

Gabriel et al. [117] trained predictive models using RF, ridge and lasso regression, and multivariable logistic regression to determine those patients who would not need prolonged hospital LOS after THR. The discriminatory ability was reported as 0.735 (95% confidence interval, 0.675–0.787) using the area under the receiver operating characteristic curve (AUC) for multivariable logistic regression (the best-fit algorithm). Also, '$P=0.37>.05$' was obtained as fitting goodness by the Hosmer–Lemeshow test. Nine variables - age, sex, anaemia, opioid use, obesity, metabolic equivalents score, chronic obstructive pulmonary disease, primary anaesthesia type, and hypertension - were included in the proposed calculator. The authors stated that this model might assist clinicians in the strategic planning of bed availability to reduce both overcrowding and underutilisation [117]. However, this sort of single-institution studies needs to add external validations and use larger sample sizes before reporting big statements.

Prediction of patient-reported outcomes (130,945 observations) by eight supervised binary classifiers (logistic regression, extreme gradient boosting, multistep adaptive elastic-net, RF, neural net, Naïve Bayes, k-Nearest Neighbors and boosted logistic regression) was the aim of a study on THR and TKR [118]. The generic and disease-specific improvement was considered as the dependent outcome based on the Oxford Hip and Knee Score (Q score) and the EQ-5D-3L visual analogue scale (VAS). Results showed that RF, extreme gradient boosting, linear model, and multistep elastic net had the highest overall J-statistic (as a statistic that shows diagnostic tests' performance). The AUC of the best-fit models was reported as around 0.86 (VAS) and 0.70 (Q score) for knee replacements, and 0.87 (VAS) and 0.78 (Q score) for hip replacements. All these models were used to depict the most significant variables but some methods, such as RF with random permutations, can introduce bias and artificial variable selection under specific circumstances [119, 120]. If several significant variables were correlated, they share the importance, suggesting that the variable importance may be shown lower than the reality [121].

2.4 Support-Vector Machine (SVM)

The SVM is a supervised ML algorithm, suitable for creating subtle patterns from complex datasets in both classification and regression problems. The SVM classifier was examined through an image-based approach for its usefulness in rating the corrosion damage on the THR prostheses (at the head-neck taper junction) [122]. The classifier was applied to capture local and textural information (as two approaches of object recognition); then, Goldberg′s scores were given to rank the images. The hyperparameters were tuned to minimise the cross-validation error by Bayesian optimisation; the features with greater discriminatory power were selected after analysis of the Neighbourhood components as a supervised learning method to classify the multivariable dataset into separate groups. An accuracy level of 85% was obtained using five-fold cross-validation, whereas a limited pool of available prostheses made a significant limitation in terms the validity. Fontana et al. [123] investigated whether ML algorithms are able to predict the patients who will attain Minimal Clinically Important Difference in THR and TKR post-operatively. Based on patient-reported outcome measures (PROMs), 6,480 TKRs and 7,239 THRs were selected from only a single hospital. Linear SVM, logistic LASSO, and RF were trained on 80% of the dataset to predict two-year minimal clinically important differences. The AUCs of the three ML methods varied from 0.60 to 0.89 with the best result for the LASSO but Theoretically these values cover a range of poor to acceptable prediction but the presence of high multicollinearity breaks one of the assumptions promising that logistic regression can produce unbiased coefficients. Although the authors noted that ML holds much guarantee for assisting as a clinical decision-making support system, it should be considered that most similar studies were only limited to small number of observations.

2.5 Naïve Bayes

Ramkumar et al. [103] aimed to develop and validate a Naïve Bayesian model using pre-operative primary THR data to predict LOS and patient-specific inpatient payments, and then recommended the use of a risk-adjusted patient-specific payment model that reflects patient comorbidity. The data of 122,334 primary THRs, including race, age, gender, and comorbidity scores, was used to train and evaluate the model using AUC and training accuracy. Inpatient payments were categorised as the output variable, and the AUC showed the validity of 0.71 and 0.87 for payment and LOS, respectively. Naïve Bayesian methods assume conditional independence which, however, fail to identify confounding variables. The validation of the developed model required that, first, an initial viability be established before proceeding with the resource-intensive task of developing other available models. This may mean that the other ML methods such as deep learning can create a more accurate model [124, 125]. SVM and NNs algorithms can take into account confounding relationships among the variables and may create better machine automation [125].

2.6 Deep Learning

There are several major differences between deep learning and other ML methods [126]. Deep learning is a subset of machine learning that reproduces the mechanisms of the human brain in learning from big data and generating patterns for decision makings. Deep learning techniques have become popular in research studies since they can automatically perform the raw data engineering by finding the optimal inner representation, which is necessary for the discriminative (mapping) task. Deep learning methods are often mysterious because of their black-box nature, which is often the main source of concern in medical applications [127]. However, they can analyse data efficiently and can capture the more complex structure of big datasets for THR and TKR despite computational complexity [128]. For instance, in a recent study, Qiu et al. [129] used a large commercial claims dataset to identify patients with a strong likelihood of requiring TKR and THR surgeries. Supervised ML methods (RF, LASSO, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)) were investigated using 540,000 observations of patient data and 2,000 variables. The deep learning methods showed much better performance than LASSO, RF, and RNN (a common type of ANN in the medical area [130, 131]) with a pooling mechanism that recorded the best accuracy using different metrics: 0.8339 ± 0.0024 as AUC and 0.0662 ± 0.0008 for precision with recall set to 0.9. As a function to reduce the number of parameters and computation in the network, the applied pooling mechanism positively influenced the performance by detecting the additional signals from the hidden intermediate states.

2.7 Other Machine Learning Methods Used in THR and TKR

In one of the first studies to develop a pre-operative algorithm for predicting post-operative opioid use after THR [132], five ML algorithms (stochastic gradient boosting, RF, SVM, NNs, elastic-net penalized logistic regression) were developed. The elastic-net penalized logistic regression attained the best performing method for calibration, discrimination (C-statistic = 0.77), and decision curve analysis; whereas, the NNs and stochastic gradient boosting models recorded the same AUC (0.77) with elastic-net penalized logistic regression.

Predictive risk models were developed and validated in another study [133] to forecast the risk of death and major complications after THR and TKR. This involved 70,569 observations of OA patients who received primary THR and TKR and the highest C-statistics and bootstrapped confidence intervals (CIs) were reported for 30-day mortality (0.73; 0.66-0.79) and cardiac complications (0.75; 0.71-0.79) based on the cross-validation of the boosted regression models. The lowest values were also reported for returns to the operating room (0.60; 0.57-0.63), and for deep vein thrombosis (0.59; 0.55-0.64).

A similar study to predict the risk of complications and 30-day mortality after TKR and THR trained LASSO regression models on 107,792 data, including clinical inputs and pre-operative demographic variables. The results demonstrated an acceptable prediction accuracy for death (0.73; 95% CI, 0.70-0.76), a renal complication (C-statistic, 0.78; 95% CI, 0.76-0.80), and a cardiac complication (0.73; 95% CI, 0.71-0.75) within 30 days of arthroplasty, although poor accuracy was reported for venous thromboembolism (C-statistic, 0.61; 95% CI, 0.60-0.62). Importantly, it was suggested that these are the most accurate and validated prediction models; however, the models performed poorly in terms of external validation (prediction of outcomes from another dataset) [134].

Given the effect of THR on health-related quality of life (HRQoL), Nemes et al. [135] suggested a clinical decision support system (DSS) using Swedish joint registry data to help clinicians evaluate the future profits of THR by providing predictions of 1-year post-operative HRQoL. Three groups of supervised ML algorithms were used: (1) linear regression and its variants, (2) nonlinear regression algorithms, and (3) classification trees and rule-based models. The multivariate adaptive regression splines (R^2 = 0.158) were shown to have the best predictive capability. However, it was not significantly better than the developed linear regression model (R^2 = 0.157). Although eleven patient-related predictors were considered, more variables need to be analysed as predictors to construct a comprehensive and successful DSS. There is no set criterion on a good R^2 value as it may increase by adding even non-important predictors in a multivariate model; it is often preferred to compare the performance of models with the same variables [136].

To predict patients′ pain and function levels after undergoing TKR, 1,649 patient-reported data in the UK were studied with the aim of training and validating a supervised ML model. Clinical factors and patient characteristics were used as pre-operative inputs to predict the Oxford Knee Score (OKS) after 12-months of TKR. This prediction model provided an individualised estimation of post-operative OKS, and also changes in OKS. The bootstrap backward linear regression showed predictive validity with R^2 of 0.175 (internal validation) and 0.211 (external validation) [137]. These low values explained 17.5% and 21.1% variability in the outcome, suggesting that the models′ generalizability is dependant on considering more potential predictive factors.

2.8 Summary

Both ML and conventional statistical methods with similarities and differences have made great strides in the support they can offer clinicians, although conventional methods have been the main statistical approach in the domain of THR and TKR to date. For example, the generation of risk-predictive models is a common approach taken for estimating the risk of an event of interest occurring in post-operative THR and TKR. In previous studies, most of these models have been developed using the conventional methods (e.g. logistic regression, Cox proportional hazards regression) [138, 139] rather than the more modern ML strategies. These strategies are becoming the main approach for addressing prediction problems across a wide range of industry and science domains. Although to date there has been very limited adoption of these strategies for the purpose of THR and TKR predictions, it is anticipated that more studies will be published on ML predictive models for THR and TKR.

One misconception is that conventional statistical methods rely on predetermined assumptions and mathematical equations to formalise relations between the variables, whereas ML techniques use the data to recognize these relationships [140]. The key benefit of ML methods over conventional statistical methods is the ability to link a large amount of data and variables together and capture complex non-linear relationships. ML, as a useful and powerful set of computational tools, is now a common choice for the development of predictive models in the medical community [141-143]. The successful adoption of several Electronic Medical Record (EMR) systems developed for various purposes (prognosis, diagnosis, or treatment) have been noted in several studies [144, 145]. Greatly improved subsets of ML models, known as ANNs, have been notable in total joint replacement contexts because of a great potential for processing "big data" [146].

2.9 Limitations of ML and Potential for Future Research

ML has proved its undeniable capability, although it is not free of issues. The accuracy of predictive models is dependent on the quality of the data sources, and predictions may be significantly affected by the amount of data and the number of variables included. Therefore, care should be taken when dealing with limited data, as it is not advisable to report that these models are reliable with acceptable accuracy levels. Furthermore, ML models should be assessed and evaluated using a randomised cohort of studies and controlled trials in real-world settings. Hence, more improvements are needed in ML orthopaedic applications to translate the research aims into clinical practices.

It is essential to understand the difference between two different types of studies with a focus on the impact of variables on the outcome or predicting outcomes for a separate data. While ML has the potential to offer more accurate predictions, this can cost a poorer understanding of the relationships among the variables. The output of ML models needs to be interpreted carefully, and the expectations of predictive analytics can be raised with a consciousness of the matters associated with misinterpretation and over-fitting in clinical settings.

To date, multivariable predictive models have been developed for THR and TKR based mainly on patient-reported factors and imaging variables. The literature shows that ML adoption for post-operative THR and TKR is still in the basic phase with only a few studies confirming that the models are entirely available for a THR or TKR practice. This suggests future research opportunities for studies on the post-operative clinical outcomes of THR and TKR. There is still a need for models that can predict various outcomes such as the early identification of prostheses outliers based on the available big data from the national joint registries around the world. Perhaps, this indicates that now is the time to enter a new era of THR and TKR by developing decision-making support systems comprising effective predictors based on big data. A future global direction of ML in the domain of joint arthroplasty could be to enable surgeons to determine what is the best for their patients.

2.10 Research Gap and Objectives

Because changes occur in the design and performance of devices over time, the first aim of this study was to identify more specific and relevant comparator groups in order to better reflect contemporary surgical practice in primary total hip and knee communities. Ideally, the early identification of outliers needs a time-to-event outcome while limiting the confounding effects of patient characteristics and device components. Given that ML is one approach that allows us to consider many variables simultaneously to reduce the impact of confounding, this research then compared the effectiveness of using either Random Survival Forest (RSF) or regularized/unregularized Cox with control for patient and associated device confounding according to current standard techniques.

Chapter 3. The Most Appropriate Comparator in Assessing the Performance of Hip Prostheses

3.1 Overview

For end-stage hip osteoarthritis (OA), total hip replacement (THR) is the surgical procedure [3, 147]. Even though joint replacement is an effective surgical procedure with high success rates, concerns continue to exist with respect to variation in prosthesis performance. In particular, prostheses introduced to the market are consequently shown to have a higher than anticipated revision rate (HTARR). Recent data show that the outcomes of THR have improved over time but suboptimal results due to less than satisfactory implant performance do still occur [5]. An important role for joint replacement registries is to monitor the comparative performance of implants to identify factors that are associated with higher rates of revision.

Registries record detailed information on procedures performed as well as patient outcomes [5, 47, 148-150]. They are also able to deliver population-based data on the comparative result within a community. Outcome data on the revision rate of individual devices are essential to allow an evidence-based method for prosthesis selection. Analyses of registry data have found that the majority of the prostheses currently in use have satisfactory outcomes [66, 151]. However, a number of prostheses have been identified as having a rate of revision that is much higher than other prostheses within the same class. The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) registry detects these as “prostheses with a higher than anticipated rate of revision”, or outliers [5].

The AOANJRR has developed a standardised multi-stage approach for detecting outliers, which includes a preliminary test based on a comparison of the revision rate of an individual prosthesis with the comparator group, defined as all other prostheses in the same procedure class. This is done by comparing the revision rates of individual prostheses to all other prostheses in that class, with the exception of large head metal-on-metal (LHMOM) prostheses. The entire class of LHMOM devices has been previously identified as having a significantly higher risk of revision. The purpose of this activity is the early identification of those devices that are not performing to standard. The identification of outlier devices assists surgeon prosthesis selection, improves patient outcomes, and reduces healthcare costs [152].

The current comparator used by the AOANJRR has some significant limitations as non-routine procedures reflected by the use of complex prostheses are not excluded.

Prostheses parts or bearing surfaces that are known to be associated with higher revision rates are currently included. High-risk prostheses such as modular-neck stems or prostheses used for specific purposes (including constrained, dual mobility, and head size smaller than 28 mm) are still considered in the current comparator [153-155]. In addition, improvements have been made in the design, manufacture and performance of many devices over time. As a consequence, the comparator does not adequately reflect contemporary practices and is likely to overestimate the revision rate [5]. The aim of this study was to assess a number of different comparator options to provide a more relevant standard for evaluating the comparative performance of new devices.

3.2 Materials and Methods

The study period was from the first year that the AOANJRR collected THR data from all Australian hospitals (January 2003) to the closure of the dataset at the end of December 2019. The study population included all patients undergoing a primary total conventional hip replacement performed for primary osteoarthritis (OA). This selection initially included 413,417 procedures. A number of specific exclusions were then undertaken to assess the impact on the cumulative percent revision of the different comparator groups. The first exclusion was LHMOM, followed by other non-modern bearing surfaces (defined as all the bearing couples except metal or ceramic heads on cross-linked polyethylene and mixed ceramic-on-ceramic), and then devices with modular neck-stem design or used for specific purposes (including constrained, dual mobility, and head size smaller than 28 mm). Lastly, all remaining HTARR prostheses previously identified by the multi-stage standardised approach were excluded (see Figure 3.1). Further analyses were also conducted to study the changes in the most common types of revision and reasons for revision, and the AOANJRR standard was employed to determine the impact of modified comparator on the number of identified outliers. A comparative analysis of revision rates between the final modified comparator group and the current was conducted in regards to studying the effect of fixation options and bearing couples. This was done by undertaking the 1st stage of the AOANJRR standardised approach. The number of procedures and revisions for each study population reported each year to the registry was also detailed.

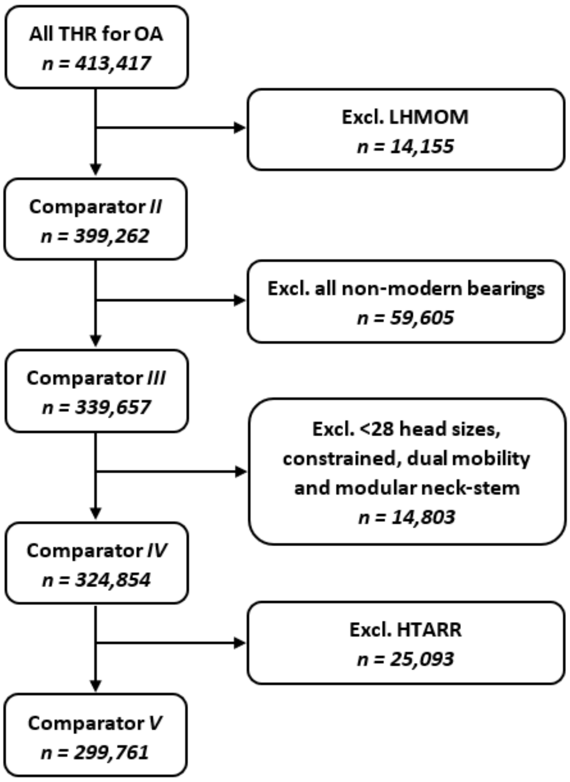


Figure 3. 1. Flow chart showing stages of exclusion criteria and study populations.
(*Note*. The AOANJRR currently uses Comparator *II* for initial screening of prostheses.)

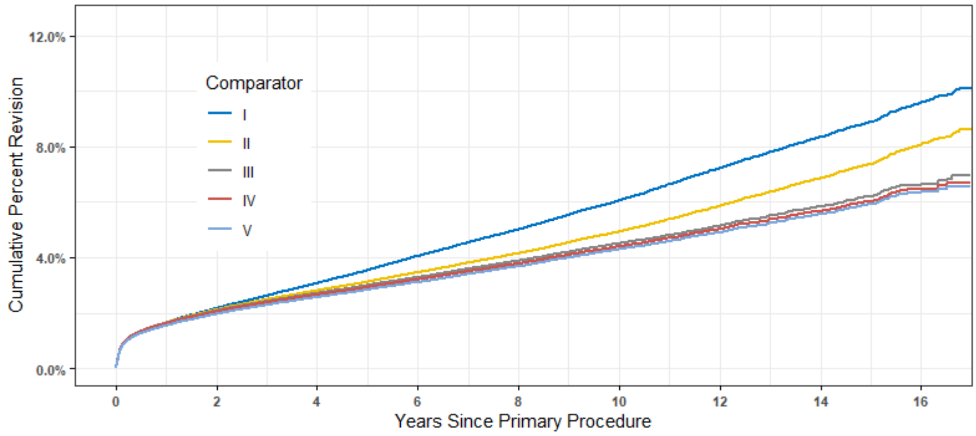
Apart from the first exclusion to generate the current comparator, all non-modern bearing surfaces were excluded because they account for less than 4% of primary THR procedures performed in 2019 [5]. Devices with modular neck-stem design or used for specific purposes were also excluded for different reasons. These components can rarely dissociate and break each other due to corrosion and wear at the modular junction [154, 155]. Dual mobility and constrained acetabular prostheses are used more frequently for patients with a higher risk of dislocation [156]. Head sizes 28 or less have a higher revision rate for dislocation and are used uncommonly in standard modern total hip [153, 157]. The remaining prostheses previously identified as having HTARR by the AOANJRR were also excluded because of the higher revision rate. The research was conducted according to the ethical principles of the Helsinki Declaration *II*. The Southern Adelaide Clinical Human Research Ethics Committee has also provided ethics approval for this study (No. 485.13).

3.3 Statistical Method

The time to first revision surgery is the outcome, and the first revision was defined as replacing, removing or adding the previous hip replacements (including one or more of the prosthetic components). Death was treated as a censored case with survival time based on the time those cases exited the study sample. Patients with no revision or death had survival times based on the time elapsed between their initial implantation date and the end of the follow-up period. Five comparator groups, the study populations (Figure 3.1), all in primary conventional THR performed for OA were studied using Kaplan-Meier (KM) survival analysis [5]. The unadjusted cumulative percentage revision (CPR) was calculated after the primary surgery, with an accompanying 95% confidence interval (CI) using unadjusted pointwise Greenwood estimates. Since each study population is defined as a subset after the exclusion, statistical comparisons of revision rates were not undertaken with Hazard Ratios (HRs). However, given the significant role of bearing surface on the survivorship of comparator, revision rates of the prostheses with non-modern against modern bearings were compared for the entire period using Cox proportional-hazards model when adjusting for age and sex. The cumulative incidence of reasons for revisions was analysed to look at the risk of most common reasons for the current and the modified comparator groups. A descriptive analysis was also performed comparatively with a view to studying the most common types of revisions. Finally, the effectiveness of the modified comparator was evaluated based on the ability to detect additional individual prostheses by performing the first two stages of the AOANJRR standard. This is done by comparing the revision rate of individual prostheses to twice the average revision rate of all other prostheses that belong to the same broad device class. In addition, the impact of confounding factors is examined by calculating age- and gender-adjusted HRs to check if there is a significant difference compared to the combined HR of the comparator group. The revision rate per 100 component years was calculated for each study population by fixation options, bearing couples, and the yearly number of procedures/revisions. The statistical analysis was performed using *R* software [158], including the packages *Survival* [159] version 3.2-11 and *Survminer* [160] version 0.4.9.

3.4 Results

Results shown in Figure 3.2 present the CPR among the comparator groups showing the proportion revised over the years. The CPR at 10 years for the current comparator was 4.93% (95% CI, 4.84–5.02), and the subsequent exclusions progressively reduced the CPR rate. The curve for Comparator *V* showed the lowest 10-year CPR of 4.30% (95% CI, 4.20–4.41). However, there was no significant difference in the 10-year CPR compared to Comparator *IV* 4.40% (95% CI, 4.30–4.50). The selection of modern bearings resulted in the greatest reduction in the comparators from CPR of 6.06% to 4.51% at 10 years (Table 3.1).



Number at risk

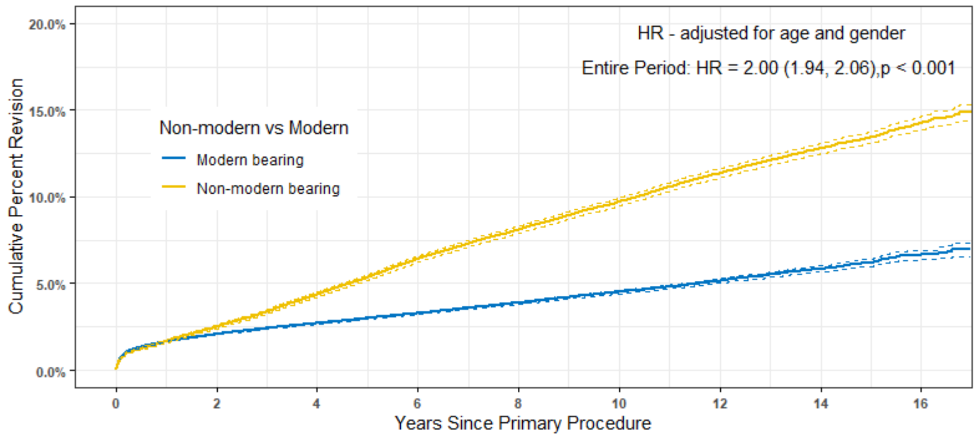
413417	329970	256019	189512	133958	87806	52329	26585	7178
399262	316524	243656	178366	123918	79537	47536	24950	6967
339657	261894	193711	133846	85394	48059	24769	11275	2620
324854	251053	184953	126980	80559	45809	23401	10376	2309
299761	233641	173510	120258	76910	44194	22579	9937	2209

Figure 3. 2. CPR of the study populations over the study period.

Table 3. 1. Yearly CPR of the study populations.

CPR	**N Revised**	**N Total**	**1 yr**	**2 yrs**	**3 yrs**	**4 yrs**
All THR	17,888	413,417	1.64 (1.60, 1.68)	2.16 (2.12, 2.21)	2.62 (2.57, 2.67)	3.07 (3.01, 3.13)
Comparator II	14,549	399,262	1.64 (1.60, 1.68)	2.12 (2.07, 2.16)	2.48 (2.43, 2.53)	2.80 (2.75, 2.86)
Comparator III	10,659	339,657	1.64 (1.60, 1.68)	2.08 (2.03, 2.13)	2.42 (2.37, 2.47)	2.71 (2.65, 2.77)
Comparator IV	9,955	324,854	1.62 (15.70, 166)	2.05 (2.00, 2.10)	2.38 (2.33, 2.44)	2.66 (2.60, 2.72)
Comparator V	8,992	299,761	1.54 (1.50, 1.58)	1.96 (1.91, 2.01)	2.29 (2.23, 2.35)	2.56 (2.50, 2.63)
	5 yr	**6 yrs**	**7 yrs**	**8 yrs**	**9 yrs**	**10 yrs**
All THR	3.54 (3.48, 3.60)	4.05 (3.98, 4.12)	4.54(4.47, 4.62)	5.01 (4.93, 5.09)	5.54 (5.45, 5.63)	6.06 (5.96, 6.16)
Comparator II	3.13 (3.07, 3.18)	3.46 (3.40, 3.52)	3.81 (3.74, 3.88)	4.14 (4.07, 4.22)	4.54 (4.45, 4.62)	4.93 (4.84, 5.02)
Comparator III	3.00 (2.93, 3.06)	3.28 (3.22, 3.35)	3.60 (3.52, 3.67)	3.88 (3.79, 3.96)	4.20 (4.11, 4.29)	4.51 (4.41,4.61)
Comparator IV	2.93 (2.87, 3.00)	3.20 (3.13, 3.27)	3.51 (3.43, 3.58)	3.78 (3.69, 3.86)	4.09 (4.00, 4.18)	4.40 (4.30,4.50)
Comparator V	2.84 (2.77, 2.90)	3.11 (3.04, 3.18)	3.41 (3.33, 3.48)	3.68 (3.59, 3.76)	3.99 (3.90, 4.09)	4.30 (4.2, 4.41)
	11 yr	**12 yrs**	**13 yrs**	**14 yrs**	**15 yrs**	**16 yrs**
All THR	6.64 (6.53, 6.75)	7.22 (7.09, 7.35)	7.81 (7.67, 7.96)	8.36 (8.20, 8.53)	8.88 (8.69, 9.07)	9.61 (9.37, 9.85)
Comparator II	5.38 (5.27, 5.48)	5.86 (5.74, 5.97)	6.37 (6.24, 6.51)	6.87 (6.71, 7.02)	7.38 (7.20, 7.56)	8.10 (7.86, 8.33)
Comparator III	4.81 (4.70, 4.93)	5.14 (5.01, 5.27)	5.52 (5.37, 5.67)	5.85 (5.67, 6.02)	6.20 (5.98, 6.41)	6.65 (6.37, 6.93)
Comparator IV	4.71 (4.59, 4.82)	5.02 (4.89, 5.15)	5.39 (5.23, 5.54)	5.69 (5.51, 5.87)	6.02 (5.80, 6.24)	6.49 (6.19, 6.78)
Comparator V	4.60 (4.49, 4.72)	4.90 (4.77, 5.04)	5.26 (5.10, 5.42)	5.57 (5.38, 5.75)	5.91 (5.69, 6.13)	6.38 (6.08, 6.68)

Given the substantial effect of excluding non-modern bearings (1st and 2nd exclusions) on the CPR of all primary THR, the risk of revision was compared by the type of bearing surfaces. Figure 3.3 illustrates a significantly higher rate of revision for the non-modern compared to the modern bearing surfaces (HR, 2.00 [95% CI, 1.94 to 2.06], p < 0.001).



Number at risk

Modern bearing	339657	261894	193711	133846	85394	48059	24769	11275	2620
Non-modern bearing	73760	68076	62308	55666	48564	39747	27560	15310	4558

Figure 3. 3. CPR of primary conventional THR by the type of bearing surface.

Table 3.2 demonstrates the additional prostheses identified by the standard using Comparator *V*. The revision rates per 100 component years of these individual devices exceeded twice that of all other total conventional hip prostheses. In addition, there are significant differences in HRs of the identified devices with the comparator *V* over the entire period. The AOANJRR standard detected 13 additional device components utilizing the modified comparator, including six femoral stems and seven acetabular components with at least 10 number of observations.

Table 3. 2. Additional device components identified utilizing the Comparator *V*.

	Descriptive information			1st stage	2nd stage	Comparator	
Acetabular cup	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs.Years**	**HR - adjusted for age and gender, *P-value***	**Current (*II*)**	***V***
Device *I*	38	384	3194.5	1.19 (0.84, 1.63)	2.53 (1.84, 3.49) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *II*	10	76	859.7	1.16 (0.56, 2.14)	2.72 (1.46, 5.07) P=0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *III*	43	712	3640.1	1.18 (0.85, 1.59)	2.12 (1.57, 2.86) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *IV*	61	453	5266.2	1.16 (0.89, 1.49)	2.67 (2.07, 3.44) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *V*	163	7006	14786.1	1.10 (0.94, 1.28)	1.23 (1.05, 1.44) P=0.008	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *VI*	72	509	6104.2	1.18 (0.92, 1.48)	2.75 (2.18, 3.48) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Femoral stem	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs.Years**	**HR - adjusted for age and gender, *P-value***	**Current (*II*)**	***V***
Device *VII*	21	184	1904.2	1.10 (0.68, 1.68)	2.47 (1.61, 3.80) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *VIII*	13	99	1184.7	1.10 (0.58, 1.88)	2.53 (1.47, 4.36) P=0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *IX*	36	344	3207.3	1.12 (0.79, 1.55)	2.51 (1.81, 3.48) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *X*	46	417	3978	1.16 (0.85, 1.54)	2.55 (1.91, 3.41) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *XI*	11	72	956.6	1.15 (0.57, 2.06)	2.68 (1.48, 4.85) P=0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *XII*	71	2232	6493.5	1.09 (0.85, 1.38)	1.52 (1.20, 1.92) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)
Device *XIII*	36	346	3136	1.15 (0.80, 1.59)	2.50 (1.80, 3.47) p<0.001	0.6 (0.59, 0.61)	0.54 (0.53, 0.55)

One of the main aims of this chapter was to report the additional individual components aligned with the current assessment of the registry in order to improve the "capture" rate of under-performing prostheses within the community. Hence, the same criteria were used for the screening test to identify prostheses with higher-than-expected revision rates than the final modified comparator (Comparator *V*). The criteria include: (1) The revision rate (per 100 component years) exceeds twice that of Comparator *V*; (2) The inclusion of the component requires at least ten primary procedures, and there have been at least two revisions; and (3) The hazard ratio of that particular prosthesis -compared to all other prostheses in the same class combined- is statistically significant through examining the impact of specific confounders (e.g. age and gender) using Cox regression models, which are known to influence implant survival and the learning curve. If an individual prosthesis was detected with at least ten procedures and two revisions using the modified comparator, and if the difference in HRs was also statistically significant, the outlier prosthesis was then listed in Table 3.2.

3.4.1 Reason for Revision and Type of Revision

Figure 3.4 details the cumulative incidence of the most common revision reasons for Comparator *V*. Figure 3.5 shows a comparative graph that provides the cumulative incidence of the same revision causes for the current comparator. The 10-year cumulative incidence with 95% CI of fracture for comparator *V* was 1.05% (1.0, 1.11), marginally lower than 1.10% (1.05, 1.15) for the current comparator. All the other common reasons for revision followed a similar pattern although the risk of revision due to loosening showed the most significant variation from 1.22% (1.17, 1.27) to 0.99% (0.93, 1.04) at 10 years. In the short term, early infection is the most probable for both study populations: 0.39% for comparator *V* and 0.38% for the current comparator at 6 months. Late loosening, as a major cause of failure in the current comparator, could be associated with the wear of hip arthroplasty components.

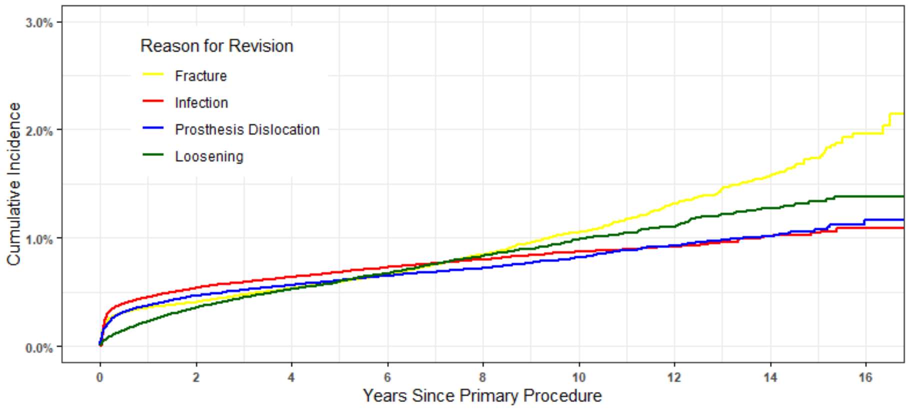


Figure 3. 4. Cumulative incidence of most common revision diagnosis for comparator *V*.

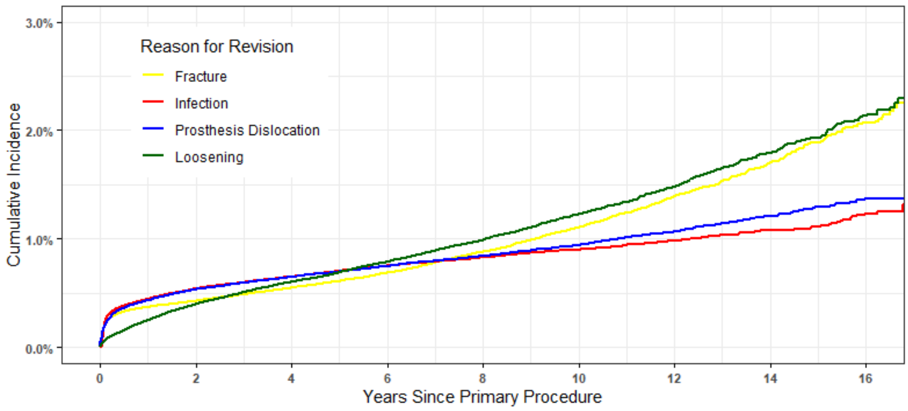


Figure 3. 5. Cumulative incidence of most common revision diagnosis for comparator *II*.

The results for types of revision (Table 3.3) show that the femoral component is the most common major revised for both comparator *V* with 35.99% (acetabular component: 17.74%, THR (Femoral/Acetabular): 9.17%, and cement spacer: 4.57%) and the current comparator with 34.28% (acetabular component: 19.93%, THR:

11.28%, and cement spacer: 4.29%). In addition, 'head/insert' had the highest percentages in the list of minor revised components followed by 'head only'.

Table 3. 3. Comparator *V* VS current comparator - Type of revision (follow-up limited to 17 years).

	Comparator *V*			Current comparator (*II*)		
Type of Revision	**Number**	**% Primaries Revised**	**% Revisions**	**Number**	**% Primaries Revised**	**% Revisions**
Femoral Component	3236	1.08	35.99	4988	1.25	34.28
Acetabular Component	1595	0.53	17.74	2900	0.73	19.93
THR (Femoral/Acetabular)	825	0.27	9.17	1641	0.41	11.28
Cement Spacer	411	0.14	4.57	625	1.16	4.29
Removal of Prostheses	51	0.02	0.57	84	0.02	0.58
Reinsertion of Components	14	0.00	0.15	15	0.00	0.10
Total Femoral	1	0.00	0.01	2	0.00	0.01
Bipolar Head and Femoral	1	0.00	0.01	1	0.00	0.01
Saddle	-	-	-	1	0.00	0.01
N Major	**6134**	**2.05**	**68.22**	**10257**	**2.57**	**70.50**
Head/Insert	2098	0.70	23.33	3022	0.76	20.77
Head Only	479	0.16	5.33	733	0.18	5.04
Minor Components	174	0.06	1.93	266	0.07	1.83
Insert Only	105	0.03	1.17	147	0.04	1.01
Bipolar Only	2	0.00	0.02	3	0.00	0.02
Head/Neck/Insert	-	-	-	68	0.02	0.47
Head/Neck	-	-	-	46	0.01	0.32
Neck Only	-	-	-	5	0.00	0.03
Cement Only	-	-	-	1	0.00	0.01
Neck/Insert	-	-	-	1	0.00	0.01
N Minor	**2858**	**0.95**	**31.78**	**4292**	**1.07**	**29.50**
N Revision	**8,992**	**3.0**	**100.00**	**14,549**	**3.64**	**100.00**
N Primary	**299,761**			**399,262**		

Note. % Primaries Revised: The proportional contribution as a percentage of all primary procedures.
% Revisions: The number of revisions as a percentage of the total number of revisions.

3.4.2 Revision Rates of Comparator Groups by Fixation

Revision rates of comparators *II* and *V* by fixation were analysed (Table 3.4) as some prostheses have more than one option for fixation. For example, a prosthesis with a recommendation to use cemented fixation may be utilised as cementless or vice-versa. Hybrid (femur cementless) has the highest rate of revision with a minimum number of observations followed by Cementless fixation for Comparator *V* and cemented for the current comparator. The best-performing fixation was Hybrid (Femur Cemented) for the final modified and the current comparator groups.

Table 3. 4. Revision rates of total hip comparator groups by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
Comparator *V*				
Cemented	182	6369	32631.5	0.56 (0.48, 0.64)
Cementless	6061	189646	1031506	0.59 (0.57, 0.60)
Hybrid (Femur Cemented)	2735	103534	586768.4	0.47 (0.45, 0.48)
Hybrid (Femur Cementless)	14	212	1036.8	1.35 (0.74, 2.26)
Current comparator				
Cemented	978	19466	32810.7	2.98 (2.80, 3.17)
Cementless	12658	263025	1106007	1.14 (1.12, 1.16)
Hybrid (Femur Cemented)	4191	130286	619730.3	0.68 (0.66, 0.70)
Hybrid (Femur Cementless)	61	640	1126.8	5.41 (4.14, 6.95)

3.4.3 Revision Rates of Comparator Groups by Bearing Surface

Table 3.5 shows revision rates for the comparator groups according to bearing surface, with a view of reporting variations in the outcomes of surface material combinations. The highest revision rates for non-modern bearing surfaces were for metal-on-metal (2.03% (1.97, 2.10)) and ceramicised metal combined with non-XLPE (1.59% (1.18, 2.10)). There are higher revision rates for ceramic and metal femoral heads combined with antioxidant XLPE within the group of modern bearings. Overall, modern surfaces indicate a lower risk of revision than non-modern bearing couples.

Table 3. 5. Revision rates of total hip comparator groups by Bearing Surface.

Bearing surface	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
Comparator *V*				
Ceramic/Ceramic	1523	55023	294751	0.52 (0.49, 0.54)
Ceramic/Metal	-	-	-	-
Ceramic/Non XLPE	-	-	-	-
Ceramic/XLPE	1792	66476	297170.8	0.60 (0.57, 0.63)
Ceramic/XLPE + Antioxidant	199	10377	26188.2	0.76 (0.66, 0.87)
Ceramicised Metal/Ceramic	-	-	-	-
Ceramicised Metal/Non XLPE	-	-	-	-
Ceramicised Metal/XLPE	706	24854	143470	0.49 (0.46, 0.53)
Ceramicised Metal/XLPE + Antioxidant	5	234	1510.5	0.33 (0.11, 0.77)
Metal/Ceramic	-	-	-	-
Metal/Metal	-	-	-	-
Metal/Non XLPE	-	-	-	-
Metal/XLPE	4677	139376	876764.2	0.53 (0.52, 0.55)
Metal/XLPE + Antioxidant	90	3421	12087.9	0.74 (0.60, 0.91)
Current comparator				
Ceramic/Ceramic	3512	90004	619411	0.57 (0.55, 0.59)
Ceramic/Metal	26	299	2792.6	0.93 (0.61, 1.36)
Ceramic/Non XLPE	321	5810	37915.4	0.85 (0.76, 0.94)
Ceramic/XLPE	2153	75873	332247.9	0.65 (0.62, 0.67)
Ceramic/XLPE + Antioxidant	281	14706	33889.2	0.83 (0.73, 0.93)
Ceramicised Metal/Ceramic	0	1	10.88	0.00 (0.00, 33.90)
Ceramicised Metal/Non XLPE	50	297	3133.2	1.59 (1.18, 2.10)
Ceramicised Metal/XLPE	719	25083	144329	0.50 (0.46, 0.53)
Ceramicised Metal/XLPE + Antioxidant	5	240	1530.8	0.33 (0.12, 0.76)
Metal/Ceramic	0	7	71.66	0.00 (0.00, 5.15)
Metal/Metal	3644	17835	179197.9	2.03 (1.97, 2.10)
Metal/Non XLPE	1615	21969	210467.7	0.77 (0.73, 0.80)
Metal/XLPE	5410	155329	977253.3	0.55 (0.54, 0.57)
Metal/XLPE + Antioxidant	143	5785	16712.2	0.85 (0.72, 1.01)

3.4.4 Number of Total and Revision by Year of Implantation

This section details the number of procedures performed and revised each year (Table 3.6) to study the difference in outcomes when only modern bearing surfaces are selected. Note that the number revised is expected to be less for the primary operations performed in later years than the number revised in earlier years as they have had less follow-up time. The use of modern bearing designs increases due to the better outcomes reported by the AOANJRR annual reports.

Table 3. 6. Number of total and revision for comparator groups by year of implantation.

Year of implant	Comparator *V*		Current comparator (*II*)	
	N Total	N Revised	N Total	N Revised
2003	4770	256	15033	1236
2004	6414	345	16125	1177
2005	7351	312	16793	1267
2006	8539	367	17408	1464
2007	9557	399	18096	1521
2008	11725	494	19790	1624
2009	14162	592	21060	1318
2010	16366	612	22565	1041
2011	18291	671	23681	961
2012	20164	614	24525	836
2013	21933	670	26337	849
2014	24130	732	28648	920
2015	25482	678	30143	829
2016	27174	709	31919	870
2017	27406	621	32617	774
2018	28130	560	33960	703
2019	28167	360	34717	498

3.5 Discussion

This study aimed to explore how the rate of revision estimated for the study populations differs according to specific prosthesis design constructs. Time to first revision was estimated for 413,417 primary total conventional hip replacements undertaken for osteoarthritis (OA) from 1st January 2003 to 31st December 2019. Survivorship analyses with stepwise exclusions were undertaken. First, large head metal-on-metal (LHMOM) procedures were excluded, followed by other non-modern bearing surfaces (defined as all the bearing couples except metal or ceramic heads on cross-linked polyethylene and mixed ceramic-on-ceramic), and then devices with modular neck-stem design or those used for specific purposes (including constrained, dual mobility, and head size smaller than 28 mm). Lastly, all remaining prostheses previously identified by the AOANJRR as having a higher than anticipated rate of revision were also excluded.

The cumulative percent revision (CPR) rate for all primary conventional THR for OA was 6.06% (95% CI 5.96, 6.16) at 10 years. After all the exclusions, the final comparator group, which only includes satisfactory-performing prostheses of contemporary design and use, had an estimated 10-year CPR of 4.3% (4.2, 4.41). This is considerably lower than the current comparator (all prostheses excluding LHMOM) used by the AOANJRR of 4.93% (4.84, 5.02). The AOANJRR standardised approach detected 13 additional device components utilizing the final modified comparator. The registry currently recommends the modern comparator for the early assessment of primary total hip prostheses.

Utilizing only contemporary device components has the potential to be a more relevant comparator for the early assessment of modern primary total hip prostheses. The use of comparator *V* led to the additional devices identified after 2nd stage of the AOANJRR standard. Increased survivorship and improved functional performance might be expected because key design modification areas are targeted to address THR complications [161]. Survivorship studies with stepwise exclusions of prostheses with high-risk designs or used for specific purposes are required periodically to introduce a more relevant and effective comparator. The modified comparator with higher sensitivity may contribute to the early detection of an outlier prosthesis with smaller sample size and shorter length of follow-up. More exclusions can potentially be added to select a group of prostheses with the lowest revision rate.

While comparator *V* had the lowest CPR, there was a slight difference between comparator *IV* and *V* (Comparator *IV* was selected for the rest of this study*).* This was because many of the HTARR prostheses previously excluded were LHMOM, modular-neck femoral components, or used as non-modern bearings [154, 162, 163]. Identification can bring a device to the attention of surgeons, alerting them to the fact that it shows a higher rate of revision [9]. However, it has become evident that the method to identify outliers may be too broad, and it is crucial to perform a careful comparative analysis of total hip prostheses. The AOANJRR approach takes into account the complexities of a small number of procedures [116], devices implanted by a single surgeon [164], and the effect of other components on the surveillance of a device [165]. After identifying a device by the AOANJRR, use of the device usually declines with a positive impact on subsequent patient outcomes. For example, ASR acetabular was first reported in 2008, then removed from the market, and the use of LHMOM prostheses was subsequently reduced [166].

There were several limitations to this research. Although the new comparator could successfully contribute to the early identification of specific prostheses within a broader group, a reduction in the number of observations available for analysis may decrease the statistical precision. More variables may produce the variance in survivorship results, such that the impact of patient characteristics was not studied on the comparator due to the study design. The other limitation includes the descriptive nature of analysing the type of revision that has not been adjusted for confounders. However, there are also some strengths, including large high-quality data with minimal loss of follow-up over a longer-term period provided by the AOANJRR, which allowed us to compare the study populations effectively. Registry outlier detection is a process that needs to evolve for optimising the detection. This would be enhanced by international collaborations between registries including data sharing [167]. Results of this study indicate that increasing the relevance of the comparator may be helpful for the early identification of a higher number of outlier prostheses.

3.6 Summary

Using a comparator that only includes contemporary devices with modern bearings and excludes special devices used in more complex primary procedures has the potential to improve the early assessment of modern primary total hip prostheses sensitively.

Chapter 4. Can Machine Learning Algorithms Contribute to the Early Identification of Primary Total Hip Outliers?

4.1 Overview

Given their extensive usage and the presence of poor-performing prostheses, total hip arthroplasty devices are among the most relevant medical devices with a lack of pre- and post-market safety assurances [3, 157]. It is known that there is variation in the safety and effectiveness of hip device components [116, 152]. While most prostheses perform acceptably, some of them may have higher than anticipated rates of revision. This variability underlines the need for attentive post-market surveillance of hip prostheses for the early detection of poor-performing components in the community [165, 168, 169]. National arthroplasty registries have acted critically in detecting these devices that are performing poorly [5, 66, 148, 149, 170, 171]. Data collected, analysed and reported by registries exposed the issue and led to the identification of prostheses with higher than anticipated revision rates called outliers.

There is growing agreement by the community that large-scale evaluations of prostheses using data from all joint registries are crucial for indicating if a device is at increased risk of revision [167, 172]. The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) has established an effective multistep approach to inform surgeons about the relative performance of prostheses [5]. Arthroplasty devices are composed of multiple components combined in a prosthesis construct to ensure the success of the procedure. Femoral stems and acetabular components are two major components, and revision surgery may mostly occur due to the failure in one or both of these total hip components. Identifying specific components that show a much higher risk of revision is challenging as numerous individual components are used in different combinations.

The initial screening effectively flags the hip components but does not account for revision rate variations over time [165]. This may cause difficulties in detecting a difference if the higher risk of revision happens later in the follow-up time [173]. The method also does not address the potential presence of other confounding factors due to device and patient variables. Ideally, an approach uses a time-to-event endpoint to identify individual components with a much higher rate of revision surgery while also reducing the confounding effects of device and patient characteristics in other components. Machine learning (ML) methods are attractive for this sort of problem because they are able to handle high-dimensional data, which conventional methods generally cannot. In addition, the methods address the additional difficulty introduced by the confounding effects. The principal objective of this study was to evaluate the use of ML methods for assessing the surveillance of total hip prostheses. The effectiveness of the methods was determined based on their ability to detect the same outliers identified by the AOANJRR gold standard.

4.2 Materials and Methods

The dataset for this research consists of 163,356 primary total conventional hip procedures with a primary diagnosis of Osteoarthritis (OA). The study period was 1st January 2015 – when the registry commenced collection of body mass index (BMI) data - to 31st December 2019. The restriction to procedures only for OA accounted for 88.2% of all surgeries over this period. There were 87 acetabular components and 126 femoral stems made by various manufacturers [5]. Patient factors and device components were the predictors and the elapsed time from primary procedure to first revision was the outcome.

Each device component was distinctly introduced with an indicator variable that showed its model name. Patient covariates comprised age, gender, BMI and American Society of Anesthesiologists (ASA) score treated as potential confounders. Gender and ASA score (less than 3 vs. greater than or equal to 3) were patient covariates with two levels; age (< 65, 65-74, and ≥75 years) and BMI (< 25, 25-29.9, and ≥ 30) were classified into three levels. Head size (≤32mm vs. >32mm) and bearing surface (modern vs non-modern) were also categorized as potential confounding, each of the variables into two ordinal groups.

Modern bearings are defined as metal or ceramic heads on cross-linked polyethylene and mixed ceramic-on-ceramic. The covariates were selected to control the impacts of relatively few patient characteristics and implant types (i.e., bearing surface, femoral head size) [111]. Missing data were only present on the patient covariates (6.35% BMI and 0.41% ASA score) handled by multiple imputations using chained equations [174]. Death was treated as a censored case with survival time up to the quit date of the study sample. Patients without experiencing a revision or death have survival times based on their initial implantations and the end of follow-up.

The effectiveness of the ML techniques was assessed to account for patient and associated device confounding factors to the AOANJRR gold standard (1st and 2nd stages). The first stage (initial screening test) is done by comparing the revision rate of individual prostheses to twice the average revision rate of all other prostheses that belong to the same broad device class. In addition, the impact of confounding factors was examined by calculating age- and gender-adjusted hazard ratios (HRs) to check if there is a significant difference compared to the combined hazard rate of the modified comparator (*IV*) developed in the previous chapter.

4.3 ML Statistical Analyses

As the concept of variable selection differs from prediction, ML models need to be trained with a careful selection of hyperparameters. Two feature selection techniques were conducted to explore the significance of inputs and find their contributions effectively in the presence of confounding effects.

For the first approach, this study employed random survival forest (RSF) as an extension of the random forest algorithm to analyse right-censored survival data [175, 176]. Large forests with a group of 2000 trees were used to reduce bias in the highly-correlated structure. Each tree of the forest was grown by repetitively performing binary splits of the AOANJRR data using the log-rank test until terminal nodes had no fewer than two revisions [177]. A random set of variables including all device components and covariates were chosen as candidates to split each parent node into two daughter nodes. It is more appropriate to develop the model such that the chance of having substantial variations between variables increases. Each tree needed to be grown deep to have as many levels as possible without limiting the node depth. Variable selection is randomized with the use of the parameter '$1 <= mtry <= P$' which was fixed at '$P/4$' [116]. The number of variables considered at each split is larger than convention ($\sqrt{P}$) because the bias in feature selection with correlated predictors can be limited by an increased number of variables considered at each split [178]. A backward selection procedure was then implemented to obtain a reduced set of informative variables by computing a new RSF with the remaining variables. A similar algorithm was suggested by Ishwaran et al. [121] and Dietrich et al. [179]. Minimal depth was used for ranking the variables [180]. In a tree, minimal depth is the distance from the tree′s root node to the node a variable is first split on. The distance of each variable is recorded based on an average taken over all trees and shorter distances denote variables with stronger effects. A threshold of 0.05 was used for permutation *P*-values to determine whether the minimal depth of a component exceeds chance [116, 181]. Given the small number of permutations implemented due to high computational cost, *P*-values adjusted based on false discovery rate (FDR) were not calculated.

The second approach was applied using a combination of ML and a well-recognized conventional regression method. A regularized model with a mixture of L1 (lasso) and L2 (ridge) penalties was developed with the aim to select a subset group of components that are most predictive of survival [182, 183]. The extent of the penalties was determined based on taking a priori value for a parameter ($\alpha = 0.5$; α ranges from 0 to 1). This is the midpoint among lasso and ridge regression called elastic-net. The parameter that specified model complexity was chosen using 10-fold cross-validation [182]. No penalty was applied to patient covariates according to a tendency to fully control the impacts of comparatively few patient characteristics (including age, gender, BMI, and ASA). The regularized Cox model does not report *P*-values because it does not test variables against null hypotheses. The selected variables by the elastic-net were entered in an unregularized Cox proportional hazards model. The reported *P*-values are based on a Wald test; the *P*-values that maintain the FDR at 0.05 [184] were also calculated using the selected variables by the elastic net. The FDR at 0.05 is much less conservative and adjusts for the more actual *P*-value distribution when 5% of all declared positive variables are genuinely negative. R statistical software was used for all analyses, glmnet [185] version 4.1-1 for Cox elastic net, and the survival package [186] version 3.2-11 for unregularized Cox regression, and randomForestSRC [174] version 2.11.0 for RSF and MICE package version 3.14.0 for multiple imputations [187].

4.4 Results

Prostheses survival for 163,356 procedures recorded by the AOANJRR were provided over the study period with the yearly number at risk (Figure 4.1 and Table 4.1). The majority of patients had an ASA score less than 3 (63.47%), were female (53.25%), had an age from 65 to 74 years (36.42%), and BMI greater than or equal to 30 kg/m² (38.86%). In the study cohort, the AOANJRR standardised approach identified three acetabular components and seven femoral stems. Note that the registry has not reported a number of these devices due to other confounding effects discussed in Sections 4.5 and 4.6 of this chapter but their continual real-time performance is monitored within a community.

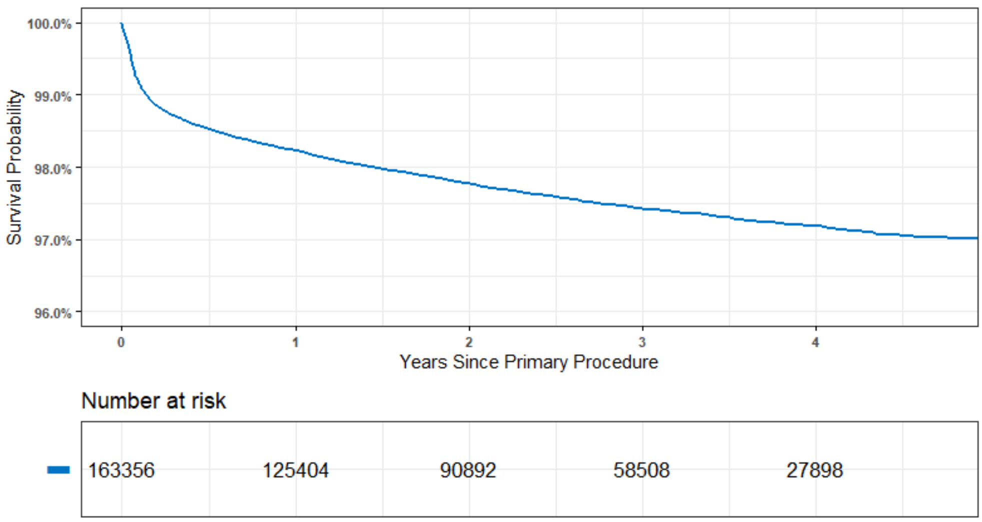


Figure 4. 1. Time to first revision for 163,356 procedures of AOANJRR data.

Table 4. 1. Individual outliers identified by the 1st and 2nd stages of the AOANJRR standard.

Component	Descriptive information			1st stage	2nd stage	
	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)	HR - adjusted for age and gender, *P*-Value	Comparator (other total)
Acetabular						
Device I	21	300	587.6	3.57 (3.29, 3.91)	3.42 (2.23, 5.26) p<0.001	0.95 (0.92, 0.98)
Device II	5	59	228.8	2.18 (2.03, 2.36)	3.14 (1.30, 7.54) p=0.01	0.95 (0.92, 0.98)
Device III	35	760	1735.6	2.02 (1.93, 2.11)	2.09 (1.50, 2.92) p<0.001	0.95 (0.92, 0.98)
Femoral stem						
Device IV	8	71	245.4	3.26 (3.01, 3.56)	4.34 (2.17, 8.68) p<0.001	0.95 (0.92, 0.98)
Device V	18	288	458.7	3.92 (3.59, 4.31)	3.28 (2.06, 5.21) p<0.001	0.95 (0.92, 0.98)
Device VI	48	1266	2271	2.11 (2.04, 2.2)	1.88 (1.42, 2.51) p<0.001	0.94 (0.91, 0.98)
Device VII	13	195	666.5	1.95 (1.86, 2.05)	2.55 (1.48, 4.40) p<0.001	0.95 (0.92, 0.98)
Device VIII	17	320	374.7	4.54 (4.25, 4.87)	3.02 (1.87, 4.86) p<0.001	0.95 (0.92, 0.98)
Device IX	28	561	1438.8	1.95 (1.86, 2.04)	2.22 (1.53, 3.22) p<0.001	0.95 (0.92, 0.98)
Device X	16	199	589	2.72 (2.54, 2.91)	3.32 (2.03, 5.42) p<0.001	0.95 (0.92, 0.98)

Note. The comparator includes all other prostheses with modern bearing surfaces excluding head sizes smaller than 28mm, constrained, dual mobility, and modular neck-stem cases. Modern bearings include only mixed ceramic/mixed ceramic and all femoral head materials used in conjunction with cross-linked polyethylene (XLPE).

The devices *IV, V,* and *VIII* were identified using both approaches and the only undetected components by ML were *II* and *VI* (Table 4.2). The random survival was able to identify eight out of ten outliers identified by the standard. These components include the acetabular components *I* and *III* and the femoral stems *IV*, *V*, *VII*, *VIII*, *IX*, and *X*. In the case of RSF, the device *X* has the lowest minimal depth rank with the most contribution to the surveillance of hip prostheses. However, Given the exact *P*-values are unknown, these ranks may not directly associate with the comparative performance of the components used.

Table 4. 2. Results for the outliers by the ML methods.

Component	Descriptive information			Random Survival Forest	Regularized/Unregularized Cox
	N Revised	N Total	Obs.Years	Minimal depth rank *Permutation P*-value	*P*-value
Acetabular					
Device I	21	300	587.6	**8** **0.019**	**-**
Device II	5	59	228.8	**20** **0.079**	**0.773**
Device III	35	760	1735.6	**15** **0.039**	**-**
Femoral stem					
Device IV	8	71	245.4	**2** **0.009**	**0.009**
Device V	18	288	458.7	**14** **0.029**	**<0.001**
Device VI	48	1266	2271	**21** **0.089**	**-**
Device VII	13	195	666.5	**13** **0.029**	**0.434**
Device VIII	17	320	374.7	**3** **0.009**	**0.012**
Device IX	28	561	1438.8	**5** **0.009**	**-**
Device X	16	199	589	**1** **0.009**	**-**

Note. Regularized Cox model selected 113 components. In the case of the regularized/unregularized Cox model approach, "-" denotes that the device was not selected; therefore, no *P*-value is provided. The Cox approach only identified one device component (*V*) when we ensured that the FDR was maintained at 0.05. In the case of the RSF, “-” denotes that the device feature was not included in any trees in the forest; therefore, no rank or *P*-value is provided.

Both the RSF and Cox techniques detected additional device components that were not previously identified by the standardised approach. A number of these devices with at least 10 observations exceeded 1.5 times the revision rate for other contemporary total hip prostheses with a significant difference in HRs (Table 4.3). The femoral stem *XIV* was detected by both the techniques and the other three were identified only by one of the approaches.

Table 4. 3. Results for the additional device components detected by ML.

Component	Descriptive information					Random Survival Forest	Regularized/Unregularized Cox
	N Revised	**N Total**	**Obs.Years**	**Revisions/100 Obs.Years (95% CI)**	**HR - adjusted for age and gender, *P*-Value**	**Minimal depth rank *Permutation P-value***	***P*-value**
Acetabular							
Device *XI*	62	1444	3466.08	1.79 (1.37, 2.29)	1.93 (1.50, 2.48) p<0.001	**4** **0.009**	-
Device *XII*	132	5048	9640.42	1.37 (1.15, 1.62)	1.26 (1.06, 1.50) p=0.008	**-**	**0.005**
Device *XIII*	40	1063	2559.11	1.56 (1.12, 2.13)	1.66 (1.22, 2.27) p=0.001	**18** **0.039**	**0.052**
Femoral stem							
Device *XIV*	14	250	804.43	1.74 (0.95, 2.92)	2.21 (1.30, 3.73) p=0.003	**17** **0.039**	**0.038**

Given a primary desire to control potential confounding, the extent of patient- and associated device confounding was evaluated. The coefficients in a Cox regression are related to HRs of device components given by the exponent of its coefficient. The next part compared the HR for specific components in two different models: (a) Cox model with a variable representing the use of that component adjusted for age and gender (2nd stage of the standard) and (b) the unregularized Cox model, which includes all the variables selected by the elastic net. This represents the effect of each component after conditioning on the selected variables (including age, gender, BMI, ASA, head size, and bearing surface). Therefore, the difference in the HRs between these two models presents the extent of potential confounding (Figure 4.2). There is at least reasonable evidence of confounding for most components; relative changes in model coefficients range from 38% for the device *V* to 204% for the *II*.

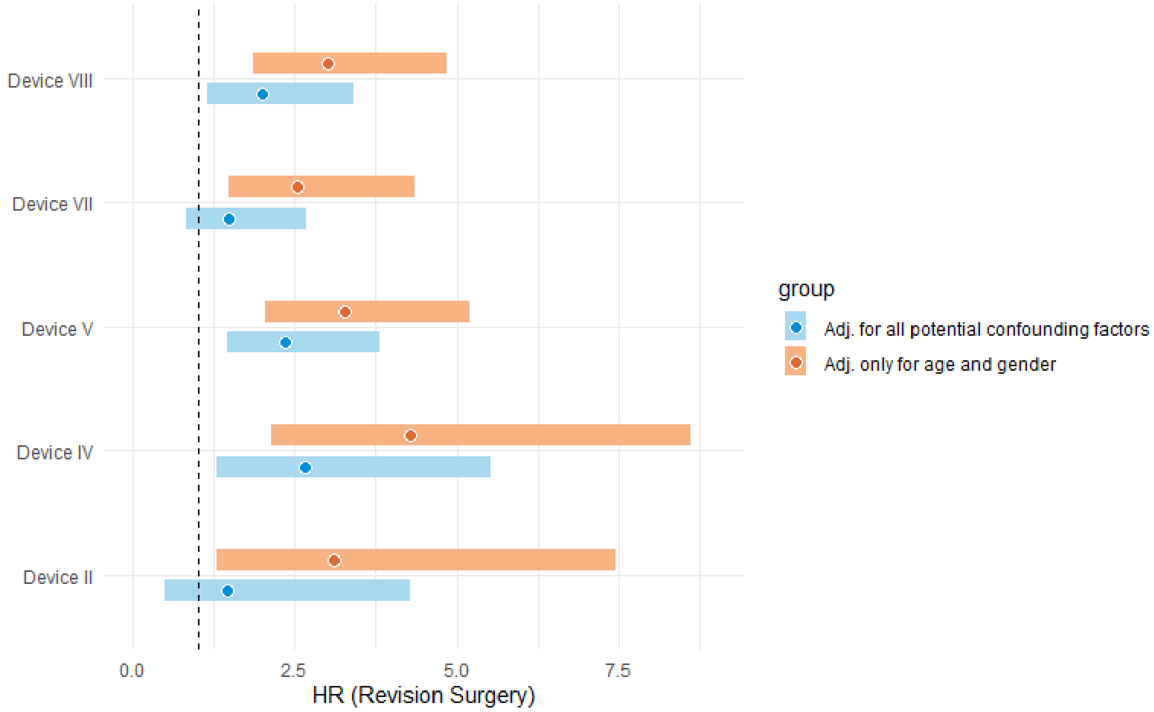


Figure 4. 2. HR comparison to illustrate the effect of potential confounding.

%Diff = [ln ($HR_{Adj.\ for\ age\ and\ gender}$) – ln ($HR_{Adj.\ for\ all\ potential\ confounding\ factors}$)]/[ln ($HR_{Adj.\ for\ all\ potential\ confounding\ factors}$)]

In 2007, the AOANJRR added a third stage that enabled an extensive examination of the prostheses identified in stage *II*. The AOANJRR professions and orthopaedic surgeons evaluate the whole range of prostheses data during a two-day workshop to finalise the list of newly-identified outliers in the AOANJRR annual reports. This stage reconsiders the potential confounding variables such as the effect of a single surgeon, catalogue ranges, or the use of a prosthesis for a specific purpose. Hence, a post-analysis was conducted to assess the interaction between surgeons and outlier subset designs (catalogue ranges).

4.5 Subsets or Catalogue Ranges

The identification method has shown an argument to carefully examine which range of prosthesis is responsible for a significantly higher than expected revision rate. The results presented in Tables 4.4 to 4.23 show that Devices *II* and *VI* might not be poor-performing prostheses as the results show an issue only with a subset of these prostheses. These devices show higher than expected revision rates for a number of subset designs while other ranges indicated acceptable outcomes. Conversely, real outliers usually have higher-than-anticipated revision rates in a significant number of catalogue ranges. Note that the number of total procedures performed with a subset design of the device needs to be reconsidered. For instance, there were significantly higher revision rates for the most subset designs of device *V* used in a greater number of procedures. Our post-analysis found strong evidence of confounding effects that may better reflect the actual performance of outlier prostheses.

Table 4. 4. Catalogue range description for Device *I* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *I*	1704607-1706607	H-A.C. CSF Plus Acetabular Cup	NO	Metal
Device *I*	1754607-1756807	H-A.C. CSF Plus Acetabular Cup	NO	Metal

Table 4. 5. Revision rates of Device *I* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
1704607-1706607	13	147	296.4	4.39 (2.33, 7.5)
1754607-1756807	8	153	291.2	2.75 (1.18, 5.41)
Total	**21**	**300**	**587.6**	**3.57 (3.29, 3.91)**

Table 4. 6. Catalogue range description for Device *II* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *II*	1800048 – 1800068	PLASMA COATED NO-HOLE/CLUSTER-HOLE ACETABULAR SHELL 48-60MM OD	NO	Metal
Device *II*	1801140 – 1801162	PLASMA COATED CLUSTER-HOLE ACETABULAR SHELL 48-62MM OD	NO	Metal

Table 4. 7. Revision rates of Device *II* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
1800048 – 1800068	3	28	110.3	2.72 (0.56, 7.95)
1801140 – 1801162	2	31	118.5	1.69 (0.2, 6.1)
Total	**5**	**59**	**228.8**	**2.18 (2.03, 2.36)**

Table 4. 8. Catalogue range description for Device *III* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *III*	012646MB-012664MB	Dual Mobility Acetabular Shell without holes	NO	Metal

Table 4. 9. Revision rates of Device *III* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
012646MB-012664MB	35	760	1735.6	2.02 (1.4, 2.8)
Total	**35**	**760**	**1735.6**	**2.02 (1.93, 2.11)**

Table 4. 10. Catalogue range description for Device *IV* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *IV*	71291101-71291901	Primary Femoral Stem Standard Offset	NO	Metal
Device *IV*	71291150 71292351	Revision Femoral Stem SH Polished	NO	Metal
Device *IV*	71290902-71290902	Sz 9 Primary Femoral Stem Standard Offset	NO	Metal
Device *IV*	71291102-71291902	Primary Femoral Stem High Offset	NO	Metal

Table 4. 11. Revision rates of Device *IV* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
71291101-71291901	4	51	180.7	2.21 (0.6, 5.67)
71291150 71292351	1	3	9.8	10.15 (0.26, 56.58)
71290902-71290902	1	1	2.5	39.66 (1.00, 220.96)
71291102-71291902	2	16	52.3	3.82 (0.46, 13.81)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 4. 12. Catalogue range description for Device *V* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *V*	4330106-4330117	Collared 133° Neck Angle	NO	Metal
Device *V*	4330206-4330214	Collarless 133° Neck Angle	NO	Metal
Device *V*	4335107-4335117	Collared 133° Neck Angle High Offset Ti6Al	NO	Metal
Device *V*	4260210-4260212	Collarless 126° Neck Angle Std Offset Ti6Al	NO	Metal
Device *V*	4265111-4265211	Collarless 126° Neck Angle High Offset Ti6Al	NO	Metal
Device *V*	4335208-4335217	Collarless 133° Neck Angle High Offset Ti6Al	NO	Metal
Device *V*	4260106-4260117	Collared 126° Neck Angle Std Offset	NO	Metal
Device *V*	4265106-4265117	Collared 126° Neck Angle High Offset Ti6Al~	NO	Metal

Table 4. 13. Revision rates of Device *V* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
4330106-4330117	10	147	214.9	4.65 (2.23, 8.56)
4330206-4330214	1	14	54.5	1.83 (0.05, 10.21)
4335107-4335117	3	67	99.7	3.01 (0.29, 4.08)
4260210-4260212	0	1	4.3	0.00 (0.00, 86.09)
4265111-4265211	0	6	13.5	0.00 (0.00, 27.23)
4335208-4335217	0	3	13.9	0.00 (0.00, 26.46)
4260106-4260117	1	24	30.7	3.26 (0.08, 18.16)
4265106-4265117	3	26	27.2	11.01 (2.27, 32.19)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Many prostheses have several catalogue ranges that are specific to particular design features. More than one catalogue range usually specifies a minor difference in the design of one specific prosthetic device. This statistical analysis was undertaken to determine the variations in these design modifications. For example, Device *IV* has the lowest revision rate with the standard design among high-offset designs and polished surfaces (Tables 4.10 and 4.11).

Table 4. 14. Catalogue range description for Device *VI* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *VI*	H49007-H49020	HA	NO	Metal
Device *VI*	H49L009-H49L020	HA Lateral	NO	Metal
Device *VI*	H49LC009-H49LC020	High Off-Set Collared	NO	Metal
Device *VI*	H49C009-H49C020	HAC collared	NO	Metal

Table 4. 15. Revision rates of Device *VI* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
H49007-H49020	9	380	729.5	1.23 (0.56, 2.34)
H49L009-H49L020	8	337	803.7	0.99 (0.43, 1.96)
H49LC009-H49LC020	16	184	214.3	7.46 (4.27, 12.12)
H49C009-H49C020	15	365	523.4	2.86 (1.6, 4.73)
Total	**48**	**1266**	**2271**	**2.11 (2.04, 2.2)**

Table 4. 16. Catalogue range description for Device *VII* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *VII*	42501006-42501017	Beaded Porous Standard Offset Reduced Neck Stem	NO	Metal
Device *VII*	42511006-42511019	Beaded Porous Lateral Offset Reduced Neck Stem	NO	Metal

Table 4. 17. Revision rates of Device *VII* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
42501006-42501017	3	32	101.17	2.96 (0.61, 8.66)
42511006-42511019	10	163	565.38	1.77 (0.85, 3.25)
Total	**13**	**195**	**666.5**	**1.95 (1.86, 2.05)**

Table 4. 18. Catalogue range description for Device *VIII* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *VIII*	0113100L - 0113108R	Anatomical Femoral Stem Ti6Al7Nb HA	NO	Metal

Table 4. 19. Revision rates of Device *VIII* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
0113100L - 0113108R	17	320	374.66	4.54 (4.25, 4.87)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Table 4. 20. Catalogue range description for Device *IX* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Type	Material
Device *IX*	523191 - 523396	Femoral Stem Standard 30-42 STD Neck	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	563514 - 563626	Femoral Stem Standard STD Neck + Lateral	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	526676 - 526684	Standard Femoral Stem 36+21 CR Neck	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	563016L - 563026R	Femoral Stem Left/Straight/Right Long 36+21 CR Neck + Lateral	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	563118L - 563216R	Femoral Stem Left/Straight/Right Long STD Neck + Lateral	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	910000001 - 910000097	Femoral Stem Standard 30-36 STD Neck	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	523418 - 523424	Femoral Stem Standard 36MM + 8 STD Neck	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	563138L - 563144R	Femoral Stem Left/Right XLong	NO	REQUIRES MODULAR BODY	Metal
Device *IX*	563158L - 563164R	Femoral Stem Left/Right XXLong	NO	REQUIRES MODULAR BODY	Metal

Table 4. 21. Revision rates of Device *IX* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
523191 - 523396	10	283	706.24	1.41 (1.32, 1.52)
563514 - 563626	9	164	436.75	2.06 (1.9, 2.25)
526676 - 526684	0	1	4.04	0.00 (0.00, 91.31)
563016L - 563026R	0	1	3.33	0.00 (0.00, 110.78)
563118L - 563216R	4	17	32.03	12.49 (8.49, 18.1)
910000001 - 910000097	0	2	6.69	0.00 (0.00, 55.14)
523418 - 523424	5	90	243.68	2.05 (1.84, 2.31)
563138L - 563144R	0	1	1.58	0.00 (0.00, 233.47)
563158L - 563164R	0	2	4.4	0.00 (0.00, 83.84)
Total	**28**	**561**	**1438.8**	**1.95 (1.29, 2.81)**

Table 4. 22. Catalogue range description for Device *X* primary total conventional hip replacement.

Model	Catalogue Range	Catalogue Description	Cement	Material
Device *X*	00787101360 - 00787101960	Femoral Stem Cemented Revision/calcar	Yes	Metal
Device *X*	00784501000 - 00784501800	Fiber Metal Midcoat Collarless Femoral Stem STD Size 10-18	NO	Metal
Device *X*	00784501230 - 00784501830	Fiber Metal Midcoat Collarless Femoral Stem Size 12-18	NO	Metal

Table 4. 23. Revision rates of Device *X* primary total conventional hip replacement by catalogue number range.

Catalogue range	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
00787101360 - 00787101960	1	17	39.7	2.52 (1.85, 3.92)
00784501000 - 00784501800	15	169	517.2	2.9 (2.71, 3.11)
00784501230 - 00784501830	0	13	31.1	0.00 (0.00, 11.86)
Total	**16**	**199**	**589**	**2.72 (1.55, 4.41)**

4.6 Number of Surgeons

The Registry is aware that a single surgeon may be responsible for a prosthesis combination with a higher-than-expected revision rate. This situation has occurred twice and, on both occasions, the use of combinations ceased after they appeared in the AOANJRR annual reports [5, 165]. Revision rates per 100 component years of the prostheses by surgeons were evaluated in order to study the effect of a single surgeon on prosthesis performance. An investigation of the number of surgeons enables clinicians to look at the other potential confounding variables contributing to the performance of a prosthesis (Table 4.24-4.33). Note that only the surgeons who used more than ten specific prostheses were included in our statistical analyses. The study was conducted to determine the performance of surgeons when they used (i) a device of interest, (ii) all the device components, and (iii) excluded the outliers detected by the AOANJRR standardised approach.

Table 4. 24. Revision rates of Device *I* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
354	Device *I*	0	21	91.2	0.00 (0.00, 4.04)
	Overall	0	86	214	0.00 (0.00, 1.72)
	Exc. outliers	0	65	122.8	0.00 (0.00, 3.01)
544	Device *I*	1	13	53.3	1.87 (0.05, 10.45)
	Overall	2	88	289.3	0.69 (0.08, 2.5)
	Exc. outliers	1	75	236	0.42 (0.01, 2.36)
587	Device *I*	3	36	68.4	4.38 (0.9, 12.81)
	Overall	12	177	506.3	2.37 (1.22, 4.14)
	Exc. outliers	3	39	50.8	5.9 (1.22, 17.26)
1246	Device *I*	2	17	26.8	7.46 (0.9, 26.95)
	Overall	7	191	406.4	1.72 (0.69, 3.55)
	Exc. outliers	5	174	379.6	1.31 (0.43, 3.07)
1357	Device *I*	1	27	105.4	0.95 (0.02, 5.29)
	Overall	12	269	620.2	1.93 (1.00, 3.38)
	Exc. outliers	9	179	451.5	1.99 (0.91, 3.78)
1726	Device *I*	1	47	32.3	3.1 (0.08, 17.26)
	Overall	9	457	1147.7	0.78 (0.36, 1.49)
	Exc. outliers	8	396	1090.8	0.73 (0.32, 1.44)
1745	Device *I*	9	113	144.5	6.23 (2.85, 11.82)
	Overall	10	145	258.4	3.87 (1.85, 7.12)
	Exc. outliers	1	28	95.6	1.04 (0.03, 5.83)

Table 4. 25. Revision rates of Device *II* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
142	*Device II*	4	57	226.9	1.76 (0.48, 4.51)
	Overall	12	237	536.9	2.23 (1.15, 3.90)
	Excl. outliers	8	180	310	2.58 (1.11, 5.08)

Table 4. 26. Revision rates of Device *III* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
155	Device *III*	0	33	80.4	0.00 (0.00, 4.58)
	Overall	3	354	914.5	0.33 (0.07, 0.96)
	Exc. outliers	3	321	834.1	0.35 (0.07, 1.05)
1041	Device *III*	15	436	1034.9	1.45 (0.81, 2.39)
	Overall	18	484	1144.4	1.57 (0.93, 2.48)
	Exc. outliers	2	42	100.2	1.99 (0.24, 7.21)
1078	Device *III*	10	96	231.5	4.32 (2.07, 7.94)
	Overall	11	103	248.7	4.42 (2.21, 7.91)
	Exc. outliers	1	7	17.2	5.81 (0.15, 32.39)
1357	Device *III*	2	63	57.8	3.46 (0.42, 12.49)
	Overall	12	269	620.2	1.93 (1.00, 3.38)
	Exc. outliers	9	179	451.5	1.99 (0.91, 3.78)
1717	Device *III*	5	60	168.7	2.96 (0.96, 6.91)
	Overall	31	1087	2799.4	1.11 (0.75, 1.57)
	Exc. outliers	26	1018	2619.3	0.99 (0.65, 1.45)

Table 4. 27. Revision rates of Device *IV* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
685	*Device IV*	4	53	188.3	2.12 (0.58, 5.44)
	Overall	8	149	323.2	2.47 (1.07, 4.88)
	Excl. outliers	4	96	134.9	2.96 (0.81, 7.59)

Table 4. 28. Revision rates of Device *V* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
587	Device *V*	1	19	38.4	2.6 (0.6, 14.5)
	Overall	12	177	506.3	2.37 (1.22, 4.14)
	Exc. outliers	3	39	50.8	5.9 (1.217, 17.26)
1246	Device *V*	2	17	26.8	7.46 (0.9, 26.95)
	Overall	7	191	406.41	1.72 (0.69, 3.55)
	Exc. outliers	5	174	379.6	1.31 (0.43, 3.07)
1357	Device *V*	1	27	105.4	0.95 (0.2, 5.29)
	Overall	12	269	620.2	1.93 (1.00, 3.38)
	Exc. outliers	9	179	451.5	1.99 (0.91, 3.78)
1421	Device *V*	0	19	9	0.00 (0.00, 41.13)
	Overall	18	490	990.8	1.82 (1.08, 2.87)
	Exc. outliers	17	445	945.7	1.80 (1.05, 2.88)
1726	Device *V*	1	61	56.9	1.76 (0.04, 9.79)
	Overall	9	457	1147.7	0.78 (0.36, 1.49)
	Exc. outliers	8	396	1090.8	0.73 (0.32, 1.44)
1745	Device *V*	9	112	144.4	6.23 (3.32, 12.74)
	Overall	10	145	258.4	3.87 (1.85, 7.12)
	Exc. outliers	1	28	95.6	1.04 (0.03, 5.83)

Table 4. 29. Revision rates of Device *VI* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
39	Device *VI*	4	11	1.9	212.05 (57.77, 542.92)
	Overall	15	394	1000.7	1.50 (0.84, 2.47)
	Exc. outliers	11	382	994	1.11 (0.55, 1.98)
153	Device *VI*	1	19	12.7	7.89 (0.2, 43.94)
	Overall	5	219	590.4	0.85 (0.27, 1.98)
	Exc. outliers	4	200	371.4	1.01 (0.29, 2.76)
156	Device *VI*	1	10	19.1	5.24 (0.13, 29.21)
	Overall	4	96	206.5	1.94 (0.53, 4.96)
	Exc. outliers	3	86	187.4	1.6 (0.33, 4.68)
275	Device *VI*	2	145	425.4	0.47 (0.06, 1.7)
	Overall	2	223	634	0.32 (0.04, 1.14)
	Exc. outliers	0	77	205.5	0.00 (0.00, 1.79)
495	Device *VI*	1	20	43.5	2.3 (0.06, 12.8)
	Overall	6	187	479.3	1.25 (0.46, 2.72)
	Exc. outliers	5	167	435.8	1.15 (0.37, 2.68)
895	Device *VI*	1	34	126.8	0.79 (0.01, 4.39)
	Overall	7	149	373.7	1.87 (0.75, 3.86)
	Exc. outliers	6	115	246.9	2.43 (0.89, 5.29)
961	Device *VI*	3	58	83.9	3.58 (0.74, 10.45)
	Overall	3	79	165.5	1.81 (0.37, 5.3)
	Exc. outliers	0	21	81.6	0.00 (0.00, 4.52)
1001	Device *VI*	1	115	347.8	0.29 (0.01, 1.6)
	Overall	11	633	1302.2	0.84 (0.42, 1.51)
	Exc. outliers	9	510	942	0.95 (0.44, 1.81)
1149	Device *VI*	6	174	190	3.16 (1.16, 6.87)
	Overall	14	392	924.2	1.51 (0.83, 2.54)
	Exc. outliers	8	217	730.8	1.09 (0.47, 2.16)
1177	Device *VI*	6	134	275.7	2.17 (0.8, 4.74)
	Overall	30	425	1303.9	2.30 (1.55, 3.28)
	Exc. outliers	24	291	1028.2	2.72 (1.49, 3.47)
1195	Device *VI*	0	36	128.8	0.00 (0.00, 2.86)
	Overall	10	752	1814.3	0.55 (0.26, 1.01)
	Exc. outliers	8	693	1675.2	0.48 (0.21, 0.94)
1218	Device *VI*	16	366	404.2	3.96 (2.26, 6.43)
	Overall	43	763	1666.6	2.58 (1.87, 3.47)
	Exc. outliers	27	397	1262.4	2.14 (1.41, 3.11)
1260	Device *VI*	1	84	85.2	1.17 (0.03, 4.33)
	Overall	4	371	772.1	0.52 (0.14, 1.33)
	Exc. outliers	3	286	686	0.44 (0.09, 1.28)

Table 4. 30. Revision rates of Device *VII* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
587	Device *VII*	8	117	413.1	1.94 (0.83, 3.81)
	Overall	12	177	506.3	2.37 (1.22, 4.14)
	Exc. outliers	3	39	50.8	5.9 (1.22, 17.26)
1066	Device *VII*	1	13	38.7	2.58 (0.06, 14.39)
	Overall	9	262	568.3	1.58 (0.72, 3.00)
	Exc. outliers	8	249	529.6	1.51 (0.65, 2.98)
1226	Device *VII*	4	63	207.1	1.93 (0.53, 4.94)
	Overall	8	107	315.2	2.54 (1.09, 5.00)
	Exc. outliers	4	44	108.1	3.70 (1.01, 9.51)

Table 4. 31. Revision rates of Device *VIII* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
804	Device *VIII*	2	93	94.2	2.12 (0.26, 7.67)
	Overall	19	1204	3078.2	0.62 (0.37, 0.96)
	Exc. outliers	17	1106	2976	0.57 (0.33, 0.91)
1041	Device *VIII*	5	110	173.7	2.88 (0.93, 6.72)
	Overall	18	484	1144.4	1.57 (0.93, 2.48)
	Exc. outliers	2	42	100.2	1.99 (0.24, 7.21)
1195	Device *XIII*	2	17	5.1	39.19 (4.75, 141.57)
	Overall	10	752	1814.3	0.55 (0.26, 1.01)
	Exc. outliers	8	693	1675.2	0.48 (0.21, 0.94)
1421	Device *XIII*	1	22	29.1	3.43 (0.09, 19.14)
	Overall	18	490	990.8	1.82 (1.08, 2.87)
	Exc. outliers	17	445	945.7	1.80 (1.05, 2.88)
1529	Device *VIII*	2	20	28.7	6.96 (0.84, 25.14)
	Overall	3	161	241.6	1.24 (0.26, 3.63)
	Exc. outliers	1	138	206	0.48 (0.01, 2.7)
1717	Device *VIII*	0	10	13	0.00 (0.00, 28.36)
	Overall	31	1087	2799.4	1.11 (0.75, 1.57)
	Exc. outliers	26	1018	2619.3	0.99 (0.65, 1.45)
1914	Device *VIII*	1	13	6	16.68 (0.42, 92.92)
	Overall	2	99	171.3	1.17 (0.14, 4.22)
	Exc. outliers	1	86	165.3	0.6 (0.01, 3.37)

Table 4. 32. Revision rates of Device *IX* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
19	Device *IX*	0	10	30.5	0.00 (0.00, 12.09)
	Overall	0	205	486.8	0.00 (0.00, 0.76)
	Exc. outliers	0	195	456.3	0.00 (0.00, 0.81)
203	Device *IX*	1	20	62.2	1.61 (0.04, 8.95)
	Overall	21	1284	3130.3	0.67 (0.41, 1.02)
	Exc. outliers	20	1264	3068.1	0.65 (0.4, 1.01)
207	Device *IX*	0	26	70.8	0.00 (0.00, 5.21)
	Overall	7	387	892.65	0.78 (0.31, 1.61)
	Exc. outliers	7	261	821.85	0.85 (0.34, 1.75)
261	Device *IX*	3	21	40.8	7.34 (1.51, 21.46)
	Overall	4	26	55.6	7.19 (1.96, 18.42)
	Exc. outliers	1	5	14.8	6.76 (0.17, 37.65)
294	Device *IX*	0	10	13.9	0.00 (0.00, 26.43)
	Overall	2	105	304.5	0.66 (0.08, 2.37)
	Exc. outliers	2	95	290.6	0.68 (0.08, 2.49)
297	Device *IX*	4	40	134.8	2.97 (0.81, 7.6)
	Overall	9	281	706.8	1.27 (0.58, 2.42)
	Exc. outliers	5	241	572	0.87 (0.28, 2.04)
562	Device *IX*	1	43	123.1	0.81 (0.02, 4.53)
	Overall	6	512	1341.1	0.45 (0.16,0.97)
	Exc. outliers	5	469	1218	0.41 (0.13, 0.96)
676	Device *IX*	0	10	36.6	0.00 (0.00, 10.1)
	Overall	0	183	457.6	0.00 (0.00, 0.81)
	Exc. outliers	0	173	421	0.00 (0.00, 0.88)
1111	Device *IX*	0	10	28.8	0.00 (0.00, 12.79)
	Overall	6	337	833.6	0.72 (0.26, 1.57)
	Exc. outliers	6	327	804.8	0.74 (0.27, 1.62)
1163	Device *IX*	0	20	49.8	0.00 (0.00, 7.4)
	Overall	6	277	733.3	0.82 (0.3, 1.78)
	Exc. outliers	6	257	683.5	0.87 (0.32, 1.91)

1372	Device *IX*	2	84	166.4	1.20 (0.14, 4.34)
	Overall	22	790	1607.5	1.37 (0.857, 2.07)
	Exc. outliers	20	706	1441.1	1.38 (0.85, 2.14)
1760	Device *IX*	0	14	19.9	0.00 (0.00, 18.49)
	Overall	5	255	524.9	0.95 (0.31, 2.22)
	Exc. outliers	5	241	505	0.99 (0.32, 2.31)

Table 4. 33. Revision rates of Device *X* primary total conventional hip replacement by surgeon id and the use of prostheses.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
747	*Device X*	15	176	524.1	2.86 (1.6, 4.72)
	Overall	21	272	725.7	2.89 (1.79, 4.42)
	Excl. outliers	6	96	201.6	2.98 (1.09, 6.48)

This study noticed important interactions between surgeons and device components, such as in the case of Devices *II*, *IV*, and *X*. Higher-than-anticipated revision rates of these devices could be associated with either the poor performance of the surgeon or the device itself or even both. The idea of assessing the number of surgeons originated from our clinical pre-knowledge intended to determine whether the performance of the prosthesis was linked to the surgeon who performed the procedure. For example, Table 4.32 gives a list of surgeons with experience using Device *IX*, showing that one poor-performing surgeon (261) had a relatively much higher revision rate than the others. It would be challenging to discuss surgeons' effects in terms of Devices *IX* and *VI* which are used by many surgeons, albeit with a variety of outcomes. However, there was no significant interaction for a number of devices such as *VII* and *VIII* because these two show a higher-than-expected rate of revisions regardless of the surgeon's performance. These findings should be considered before reporting the hip outlier prostheses.

4.7 Discussion

Early identification of outliers ideally uses a time-to-event outcome while reducing the confounding effects of other components in the device and patient characteristics. ML, which contains self-learning algorithms, is one approach to consider many variables simultaneously to limit the impact of confounding. The principal objective of this study was to compare the effectiveness of using either RSF or regularized/unregularized Cox regression to account for patient and associated device confounding factors to current standard techniques.

This study evaluated RSF and regularized/unregularized Cox regression using data from the AOANJRR to detect outlier devices among 213 individual primary total hip components performed in 163,356 procedures from 1st January 2015 to the end of 2019. Patient characteristics and device components were the inputs, and time to first revision operation was the primary outcome treated as a censored case for death. The effectiveness of the ML approaches was assessed based on the ability to detect the same outliers identified by the AOANJRR standardised approach.

The standard AOANJRR approach identified three acetabular components and seven femoral stems. The ML approaches identified some but not all the outliers detected by the AOANJRR. Both the methods identified three of the same femoral stems, and the RSF identified the other five components, including two of the same acetabular cups and three of the same femoral stems. In addition, both the RSF and Cox techniques detected a number of additional device components that were not previously identified by the standardised approach.

This study showed that the RSF technique was more comparable to the AOANJRR standard in terms of detecting more outlier prostheses. Of the ten outliers identified by the AOANJRR gold standard, ML was able to identify eight of the same device components, including two acetabular cups and six femoral stems. The group of detected prostheses by both the feature selection techniques includes *IV*, *V*, and *VIII*. By contrast, two out of the ten listed components (*II* and *VI*) and were identified neither by RSF nor Cox. The outcome highlights the significance of studying potential confounding effects on the comparative performance of primary total hip prostheses.

The results indicate that the ML methods explored can be effective at detecting outliers. However, a single model may not necessarily be the best choice because the inclusion or exclusion of inputs can affect the strength and even sign of a given predictor. For tree growing, RSF uses random subsets of variables per node that may cause an independent split of correlated variables. This may lead to breaking the structure of highly-correlated predictors and providing an interesting approach for explorative variable-selection studies [188]. However, false-positive discoveries due to overfitting are considered to be a major problem [189]. On the other hand, the Cox regression has a significant advantage in computational cost, interpreting variable strength, and documenting confounding effects.

Feature selection may be able to offer a supplementary identification approach with the potential to identify most of the devices detected by the standard. This similarity in the results becomes more apparent by looking at the outliers reported after meeting all the stages due to further investigation of confounding factors. The AOANJRR has not reported the two non-detected devices (*II* and *VI*). However, the three identified components identified by both the techniques were detected considering larger sample sizes and over longer times [5]. These identified femoral stems include Emperion, Furlong Evolution, and MiniMax total conventional hip prostheses. The current technique used by the registry is pragmatic and effective at detecting the relative performance of total hip prostheses with a higher risk of revision through in-depth knowledge of potential confounding factors.

The current study has several limitations. The effectiveness of screening tests depends on recognizing relevant component characteristics; the process will be compromised if some attributes that contribute to the prosthesis survival are not accounted for (see Appendix A). This study included well-known clinically relevant attributes, and head size showed the most significant contribution to the initial screening of total hip devices. However, other factors related to surgeons and catalogue ranges can also be investigated. In some cases, it appeared (solely from the overall revision rate) that surgeon-specific factors contributed to a higher-than-expected revision rate. This draws attention to the need for action to be taken in regard to the impact of the surgeon and surgical procedure on the performance of prostheses. As a limitation, this thesis did not investigate key factors associated with surgeons, such as their experience and surgical volume, due to the complexity of translating this information into classified meaningful inputs.

The contrary may be a concern as well; considering too many attributes may cause delayed detection. One possibility to address this issue is to expand the dataset by involving several registries worldwide that have information on the same prostheses. The proposed methods can be applied to knee and shoulder arthroplasty devices as a research opportunity. Utilizing prediction to understand the variables linked with the outcome may improve shared decision-making, leading to fewer patients at risk of receiving poor devices.

4.8 Summary

Machine learning may be able to offer a supplementary approach to enhance the early identification of outlier devices within the community. This study showed that the RSF technique was more comparable to the AOANJRR standardised approach and head size was the most significant device-related covariate for the initial screening of total hip devices. Further studies are required to better understand the potential of feature selection techniques to improve the early assessment of total hip outlier prostheses.

Chapter 5. The Most Appropriate Comparator in Assessing the Performance of Knee Prostheses

5.1 Overview

Knee replacement was first widely performed in the 1970s and 1980s [190]. Osteoarthritis (OA) is the most common primary diagnosis for this cost-effective surgical procedure [47]. The demand for knee replacement surgery is projected to rise due to the increasingly ageing population. Early detection mechanisms are required to identify poor-performing prostheses (outlier prostheses) with unreliable clinical outcomes for patients. The identification and documentation of outlier prostheses reduce their usage leading to better clinical outcomes [116].

Joint registries aim to reduce the revision rates of arthroplasty surgeries by early detection of outlier joint arthroplasty devices [5, 34, 47]. They deliver population-based data on the comparative surveillance of prostheses within the community. Survival outcome data are essential for an evidence-based approach to identify prostheses with statistically higher than anticipated revision rates. Given the signal detection efforts to exclude outliers over time, the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) reports 90% survival for primary total knee replacement (TKR) for OA at 19 years [5].

There are different types of knee replacements that can be classified based on the type of articulation. All registries report variations in the outcomes of total knee prostheses by stability. Stability is used for various purposes and the type of stability used for prostheses may affect the overall outcomes within the same class. Most total knee prostheses implanted are either cruciate retaining (CR) or posterior stabilised (PS) prostheses [4]. In Australia, these two stability types have remained the most widely and commonly used primary TKR procedures [34, 40, 66]. On the other hand, complex designs (i.e., FS and hinged implants) are also used in a limited number of primary procedures based on the clinical circumstance [5].

The AOANJRR has developed a standardised multi-stage approach to identify primary total knee outliers by performing an initial screening test. This is done by comparing the revision rate of individual prostheses to the average revision rate of all prostheses belonging to the same broad TKR class. However, the current comparator does not adequately differentiate between complex and conventional procedures. This may result in less conventional and more complex devices being identified as being at risk [5, 66]. Given the higher associated risk of failure with complex knee prostheses, this study aims to identify more specific and relevant comparator groups by stability design to better reflect the corresponding type of prostheses.

5.2 Materials and Methods

The study period was from the first year that the AOANJRR collected TKR data from all Australian hospitals (January 2003) to the closure of the dataset at the end of December 2019. The study population included all patients undergoing a primary TKR performed for primary OA. The outcome was time to first revision surgery, defined as reoperations of previous knee replacements where one or more prosthetic device components are replaced, removed, or added. Death was treated as a censored case with survival time based on the date those cases exited the study. Patients with no revision or death had survival times based on the time elapsed between the initial surgery and the end of follow-up. Further analyses were conducted to study the changes in the most common types of revision and reasons for revision. The AOANJRR standardised approach was then employed to determine the impact of modified comparator groups on the number of identified outliers. The revision rate of primary total knee surgery by fixation, bearing surface, bearing mobility, and use of patella was also calculated to evaluate the differences between the complex and conventional study populations.

5.3 Standard Designs

The AOANJRR defines CR prostheses with a flat or bowl-shaped tibial articulation, regardless of congruency. PS design prostheses mostly offer additional posterior stability with a box and peg design; or less often using a groove and cam. The use of CR prostheses has continued comparatively constant over the last 10 years. In 2019, CR stability accounted for 71.6% of primary procedures. However, the use of PS design prostheses experienced a reduction in trend from 32.9% in 2008 to 19.2% in 2019 [5].

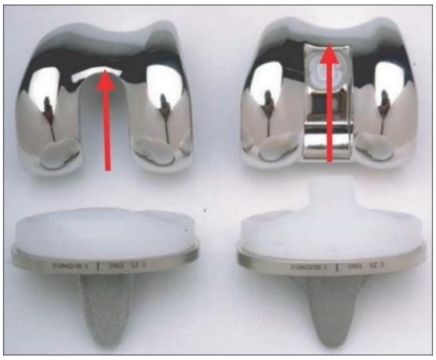


Figure 5. 1. Left presents the photograph of a cruciate-retaining and right offers posterior stabilized femoral components, showing a higher transition height from the trochlear groove to the intercondylar box in the posterior stabilized femoral component [191].

5.4 Complex Designs

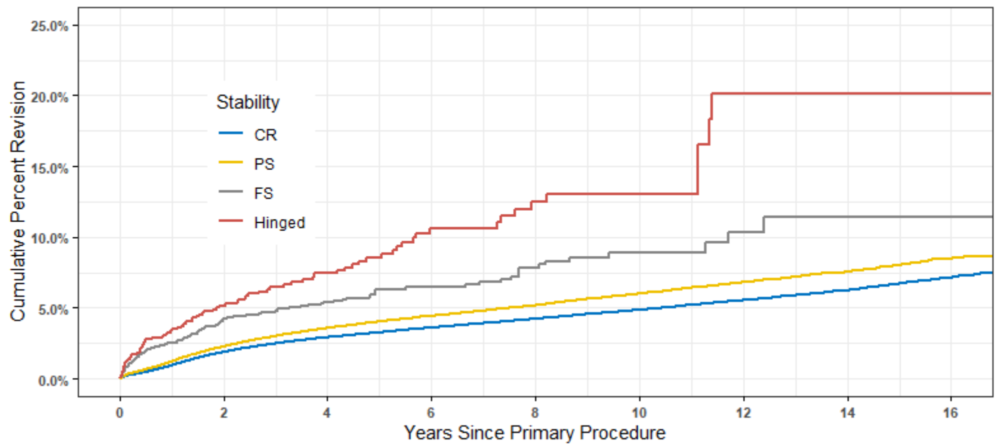
Hinged knees with added collateral and posterior ligament stability, and FS with a large peg and box design are used less often. These designs are usually considered as revision components or only performed in complex clinical situations of primary surgeries. Complex design prostheses are mostly used for patients with the primary diagnosis of tumours, fractures, and rheumatoid arthritis [5].

5.5 Statistical Methods

Two study populations in primary TKR performed for OA were studied using Kaplan-Meier survival analysis [5]. The unadjusted CPR was estimated after the primary surgery. This measure was calculated using unadjusted pointwise Greenwood estimates with an accompanying 95% confidence interval (CI). To compare revision rates between the two modified comparator groups, age- and gender-adjusted hazard ratios (HRs) for the entire period were calculated using the Cox proportional hazard model. The secondary outcome measure was the cumulative incidence of reasons for revisions. This concept was analysed to study the competing risk of most common revision reasons amongst the complex and conventional TKRs. A descriptive analysis was also performed comparatively to analyse the contribution of each reason for revision and type of revision. In addition, the CPR rate by fixation, bearing surface, bearing mobility, and the use of patella were calculated. The effectiveness of the modified comparator to detect individual prostheses was evaluated by performing the first two stages of the AOANJRR standardised approach. This involved comparing the revision rate of individual prostheses to twice the average revision rate of all prostheses belonging to the same broad device class. The impact of confounding factors was examined by calculating age- and gender-adjusted HRs to check for a significant difference compared to the combined HR of the comparator group. The statistical analysis was performed using *R* software [158], including the packages *Survival* [159] version 3.2-11 and *Survminer* [160] version 0.4.9.

5.6 Results

Fully stabilised and hinged designs show higher CPR rates than CR and PS over the entire period (Figure 5.2). Table 5.1 shows the yearly CPRs of primary TKR by stability design. The use of PS design led to a higher overall CPR than the CR design for conventional prostheses, and the hinged design had a higher CPR than FS for complex prosthesis constructs. In total, there was a higher risk of revision for the two complex designs compared to the conventional prostheses.



Number at risk

CR	463863	371776	290055	216691	152699	99904	59975	30425	8056
PS	172530	144816	114903	85841	60467	37653	19667	7708	1603
FS	2519	1794	1165	716	410	218	98	45	8
Hinged	1133	716	443	275	161	73	33	15	4

Figure 5. 2. CPR of primary total knee replacement by stability.

Table 5. 1. Yearly CPR of primary total knee replacement by stability.

CPR	N Revised	N Total	1 yr	2 yrs	3 yrs	4 yrs
CR	16,406	463,863	0.9 (0.9, 1.0)	1.9 (1.8, 1.9)	2.5 (2.4, 2.5)	2.9 (2.9, 3.0)
PS	7,725	172,530	1.2 (1.2, 1.3)	2.2 (2.2, 2.3)	3.0 (2.9, 3.1)	3.6 (3.5, 3.7)
FS	139	2,519	2.6 (1.9, 3.2)	4.2 (3.4, 5.1)	4.8 (3.9, 5.7)	5.4 (4.4, 6.4)
Hinged	86	1,133	3.5 (2.4, 4.6)	5.3 (3.9, 6.7)	6.5 (4.8, 8.1)	7.7 (5.8, 9.4)
	5 yr	**6 yrs**	**7 yrs**	**8 yrs**	**9 yrs**	**10 yrs**
CR	3.3 (3.2, 3.3)	3.6 (3.5, 3.7)	3.9 (3.8, 4.0)	4.2 (4.1, 4.3)	4.5 (4.5, 4.6)	4.8 (4.8, 4.9)
PS	4.0 (3.9, 4.1)	4.4 (4.3, 4.5)	4.8 (4.7, 4.9)	5.2 (5.1, 5.3)	5.6 (5.5, 5.7)	6.0 (5.9, 6.1)
FS	6.3 (5.1, 7.4)	6.5 (5.3, 7.7)	6.8 (5.6, 8.1)	8.1 (6.5, 9.6)	8.6 (6.8, 10.2)	8.9 (7.0, 1.07)
Hinged	8.8 (6.7, 10.9)	10.6 (8.1, 13.1)	11.0 (8.4, 13.6)	12.5 (9.3, 15.5)	13.5 (9.8, 17.2)	13.5 (9.8, 17.2)
	11 yr	**12 yrs**	**13 yrs**	**14 yrs**	**15 yrs**	**16 yrs**
CR	5.2 (5.1, 5.3)	5.5 (5.4, 5.6)	5.9 (5.8, 6.0)	6.2 (6.1, 6.4)	6.7 (6.6, 6.9)	7.1 (6.9, 7.3)
PS	6.4 (6.3, 6.6)	6.8 (6.6, 7.0)	7.2 (7.0, 7.4)	7.5 (7.3, 7.8)	8.0 (7.7, 8.3)	8.5 (8.1, 8.9)
FS	8.9 (7.0, 1.07)	10.6 (7.8, 13.7)	11.4 (8.0, 14.7)	11.4 (8.0, 14.7)	11.4 (8.0, 14.7)	-
Hinged	13.5 (9.8, 17.2)	13.5 (9.8, 17.2)	13.5 (9.8, 17.2)	13.5 (9.8, 17.2)	13.5 (9.8, 17.2)	13.5 (9.8, 17.2)

Figure 5.3 presents the CPR among the study populations for conventional and complex procedures showing the proportion revised. The conventional curve shows a 10-year CPR of 10.3% (8.6, 12.0) for the complex designs and a 10-year CPR of 5.2% (5.1, 5.2) for the conventional prostheses performed in primary TKR (Table 5.2).

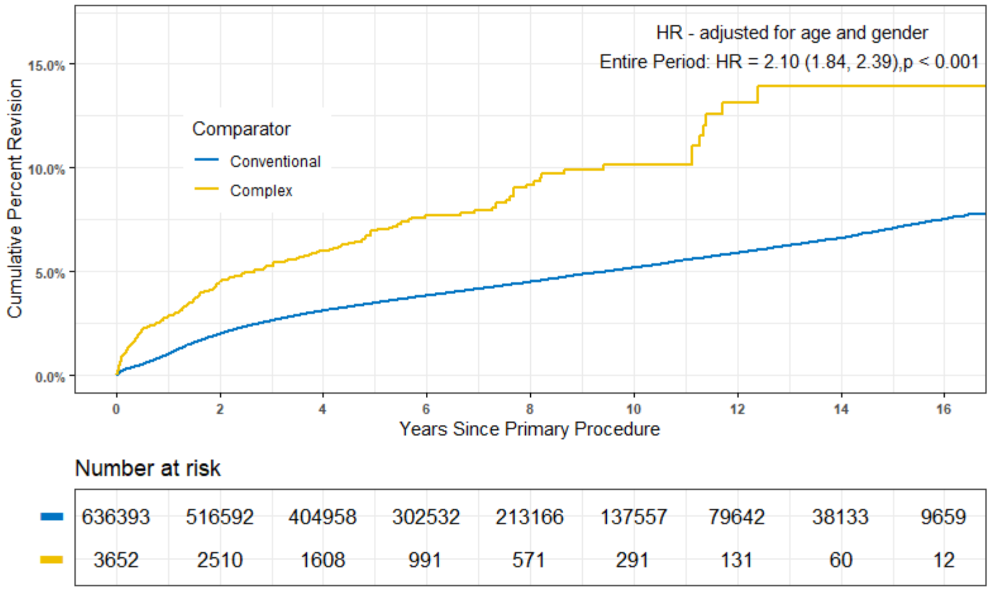


Figure 5. 3. CPR of conventional and complex comparator groups.

Table 5. 2. Yearly CPR of the comparator groups.

CPR	N Revised	N Total	1 yr	2 yrs	3 yrs	4 yrs
Conventional	24,131	636,393	1.0 (1.0, 1.1)	2.0 (1.9, 2.0)	2.6 (2.6, 2.7)	3.10 (3.0, 3.1)
Complex	225	3,652	2.8 (2.3, 3.4)	4.5 (3.8, 5.2)	5.3 (4.5, 6.1)	6.0 (5.1, 6.9)
	5 yr	**6 yrs**	**7 yrs**	**8 yrs**	**9 yrs**	**10 yrs**
Conventional	3.5 (3.4, 3.5)	3.8 (3.8, 3.9)	4.2 (4.1, 4.2)	4.5 (4.4, 4.5)	4.9 (4.8, 4.9)	5.2 (5.1, 5.3)
Complex	7.0 (6.0, 8.0)	7.7 (6.6, 8.8)	7.9 (6.8, 9.0)	9.2 (7.8, 10.6)	9.9 (8.3, 11.4)	10.3 (8.6, 12.0)
	11 yr	**12 yrs**	**13 yrs**	**14 yrs**	**15 yrs**	**16 yrs**
Conventional	5.5 (5.5, 5.6)	5.9 (5.8, 6.0)	6.2 (6.1, 6.3)	6.6 (6.5, 6.7)	7.1 (6.9, 7.2)	7.5 (7.4, 7.7)
Complex	10.3 (8.6, 12.0)	13.1 (10.2, 16.0)	13.9 (10.6, 17.0)	13.9 (10.6, 17.0)	13.9 (10.6, 17.0)	13.9 (10.6, 17.0)

Our findings show a higher sensitivity obtained for the early assessment of conventional prostheses with the potential to detect outliers with greater accuracy. The modified conventional comparator caused the identification of additional conventional and fewer complex prostheses through stages *I* and *II* of the standardised approach. The non-detected devices with complex designs could not be actual outlier prostheses. They were mostly used in high-risk surgeries and needed to be compared statistically to the more relevant comparator.

Table 5.3 shows the details of the two additional combinations detected utilising the modified comparator focusing on the routinely used devices. Defining the specific comparator for complex design prostheses reduced the number of identified outliers by the AOANJRR standard (Table 5.4). The revision rates of these devices exceeded stage *I* but there was no significant difference between the HRs of the listed components and the complex comparator. The use of modified comparator groups caused a meaningful change in the number of identified prostheses as being at risk.

Table 5. 3. Additional identified conventional prostheses using the modified comparator.

	Descriptive information			1st stage	2nd stage	Comparator	
Femoral/Tibial	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)	HR - adjusted for age and gender, *P-value*	Current	Conventioanl
Device *I*	43	481	3555.7	1.21 (0.87, 1.63)	2.17 (1.61, 2.93) p<0.001	0.61 (0.6, 0.61)	0.60 (0.59, 0.61)
Device *II*	58	438	4844.4	1.20 (0.91, 1.55)	2.37 (1.83,3.06) p<0.001	0.61 (0.6, 0.61)	0.60 (0.59, 0.61)

Table 5. 4. Non-detected complex prostheses using the modified comparator.

	Descriptive information			1st stage	2nd stage	Comparator	
Femoral/Tibial	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)	HR - adjusted for age and gender, *P-value*	Current	Complex
Device *III*	11	124	655.6	1.68 (0.84, 3.0)	1.18 (0.64, 2.16) P=0.594	0.61 (0.6, 0.61)	1.42 (1.23, 1.61)
Device *IV*	21	211	974.4	2.15 (1.33, 3.29)	1.44 (0.92, 2.26) p=0.108	0.61 (0.6, 0.61)	1.42 (1.19, 1.58)
Device *V*	27	478	2121.7	1.27 (0.84, 1.85)	0.92 (0.62, 1.38) p=0.694	0.61 (0.6, 0.61)	1.42 (1.25, 1.66)
Device *VI*	7	124	476	1.47 (0.59, 3.03)	0.85 (0.40, 1.82) p=0.685	0.61 (0.6, 0.61)	1.42 (1.24, 1.62)
Device *VII*	3	38	231.8	1.29 (0.27, 3.78)	0.96 (0.31, 3.01) p=0.947	0.61 (0.6, 0.61)	1.42 (1.24, 1.62)
Device *VIII*	8	115	371.3	2.15 (0.93, 4.24)	1.31 (0.64, 2.65) p=0.456	0.61 (0.6, 0.61)	1.42 (1.22, 1.60)
Device *IX*	17	295	1074.1	1.58 (0.92, 2.53)	1.04 (0.63, 1.71) p=0.874	0.61 (0.6, 0.61)	1.42 (1.22, 1.62)
Device *X*	6	75	433.6	1.38 (0.51, 3.01)	0.85 (0.38, 1.92) p=0.701	0.61 (0.6, 0.61)	1.42 (1.24, 1.62)

5.6.1 Reason for Revision and Type of Revision

The descriptive results of reasons for revisions are listed in Table 5.5 for the two modified comparator groups. For both the comparator groups, infection was the most common reason for revision. Infection occurred in 24.2% (5846 of 24131) of conventional TKR prostheses, followed by loosening (23.9%), patellofemoral pain (9.2%), and instability (8.2%). Infection occurred in 52% (117 of 225) of complex knee procedures, followed by fracture (9.8%), loosening (9.3%), and instability (6.7%).

Table 5. 5. Comparator groups - Reason for revision (follow-up limited to 17 years).

	Complex			Conventional		
Reason for Revision	**Number**	**% Primaries Revised**	**% Revisions**	**Number**	**% Primaries Revised**	**% Revisions**
Infection	117	3.2	52.0	5846	0.9	24.2
Fracture	22	0.6	9.8	750	0.1	3.1
Loosening	21	0.6	9.3	5770	0.9	23.9
Instability	15	0.4	6.7	2108	0.3	8.2
Patella Erosion	8	0.2	3.5	1518	0.2	6.3
Pain	7	0.2	3.1	1983	0.3	8.2
Bearing Dislocation	7	0.2	3.1	148	0.0	0.6
Malalignment	4	0.1	1.8	520	0.1	2.1
Implant Breakage Tibial Insert	4	0.1	1.8	129	0.0	0.5
Incorrect Sizing	4	0.1	1.8	269	0.0	1.1
Patellofemoral Pain	3	0.1	1.3	2227	0.3	9.2
Patella Maltracking	2	0.1	0.9	161	0.0	0.7
Prosthesis Dislocation	2	0.1	0.9	66	0.0	0.3
Implant Breakage Femoral	2	0.1	0.9	24	0.0	0.1
Lysis	1	0.0	0.4	402	0.1	1.7
Implant Breakage Tibial	1	0.0	0.4	47	0.0	0.2
Heterotopic Bone	1	0.0	0.4	8	0.0	0.0
Arthrofibrosis	-	-	-	896	0.1	3.7
Wear Tibial Insert	-	-	-	366	0.1	1.5
Metal Related Pathology	-	-	-	339	0.1	1.4
Implant Breakage Patella	-	-	-	125	0.0	0.5
Synovitis	-	-	-	78	0.0	0.3
Osteonecrosis	-	-	-	51	0.0	0.2
Wear Patella	-	-	-	33	0.0	0.1
Tumour	-	-	-	19	0.0	0.1
Wear Tibial	-	-	-	9	0.0	0.0
Progression Of Disease	-	-	-	4	0.0	0.0
Wear Femoral	-	-	-	3	0.0	0.0
Incorrect Side	-	-	-	1	0.0	0.0
Post Operative Haematoma	-	-	-	1	0.0	0.0
Patella Dislocation	-	-	-	-	-	-
Other	4	0.1	1.8	230	0.0	0.9
N Revision	**225**	**6.2**	**100.0**	**24,131**	**3.8**	**100.0**
N Primary	**3,652**			**636,393**		

Note. % Primaries Revised: The contribution of each revision diagnosis as a percent of all primary practices.
% Revisions: The percentage of each revision diagnosis of the total number of revisions.

Figure 5.4 details the cumulative incidence of the most common revision reasons for complex design prostheses in primary total knee surgeries. Figure 5.5 illustrates a comparative graph that provides the cumulative incidence of the same revision causes for the conventional comparator group. The 10-year cumulative incidence with 95% CI of infection for the complex group was 4.8%, higher than the 1.1% incidence for the conventional designs. The overall risk of other revision causes for the complex designs was also higher than that of the conventional prostheses. Early infection is the most probable scenario for each study population, particularly underlined for the complex devices with 6-month cumulative incidences of 1.4%.

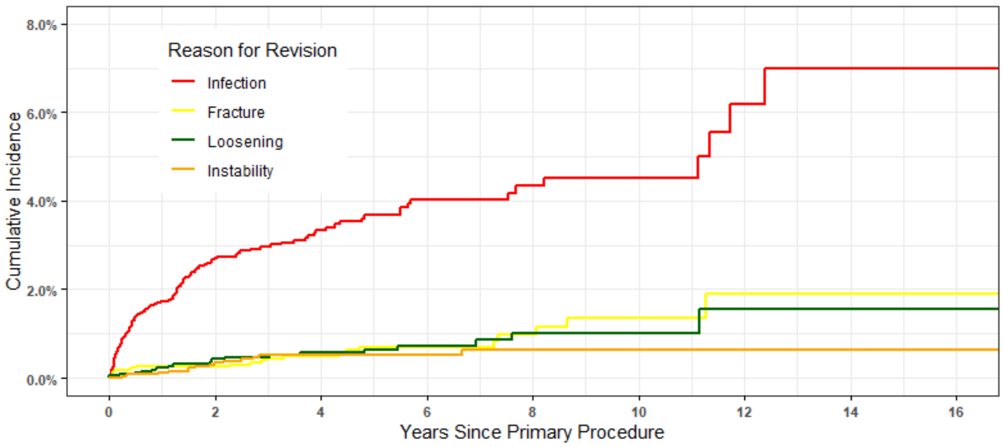


Figure 5. 4. Cumulative incidence revision diagnosis for the complex primary total knee.

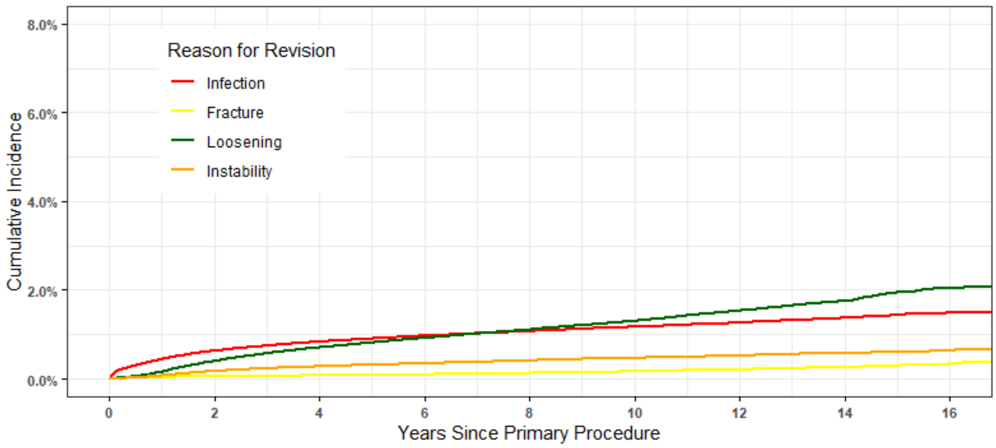


Figure 5. 5. Cumulative incidence revision diagnosis for the conventional primary total knee.

The changes of the most common types of revisions were studied by undertaking a descriptive analysis limited to 17 years of follow-up for the two study populations. Table 5.6 shows that 'TKR (tibial/femoral)' and 'insert only' were the most common major and minor types of revision respectively, for both the comparator groups. However, there are variations in the incidence probability of the other types for the complex and conventional total knee. Overall, the results for minor components (54.5%) presented a higher proportion revised than the major device components (45.5%).

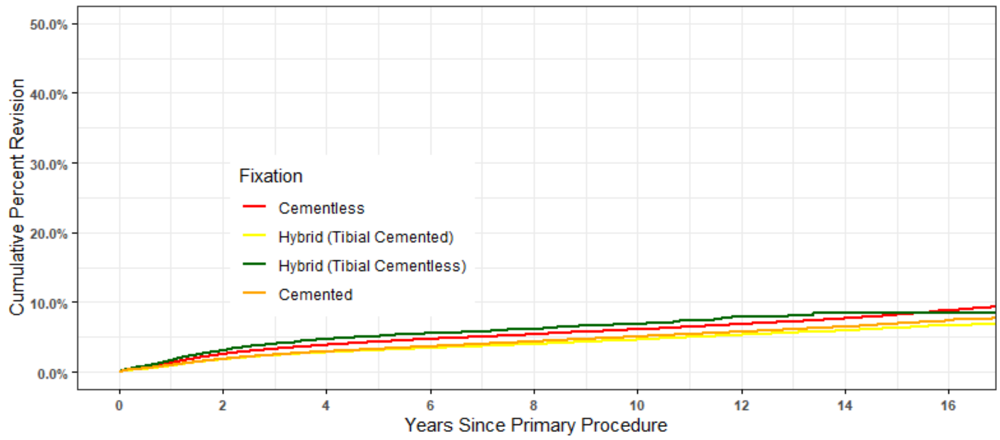
Table 5. 6. Comparator groups - Type of revision (follow-up limited to 17 years).

	Complex			Conventioanl		
Type of Revision	**Number**	**% Primaries Revised**	**% Revisions**	**Number**	**% Primaries Revised**	**% Revisions**
TKR (Tibial/Femoral)	32	0.9	14.2	6206	1.0	25.7
Tibial Component	10	0.3	4.4	2070	0.3	8.6
Cement Spacer	21	0.6	9.3	1309	0.2	5.4
Femoral Component	23	0.6	10.2	1262	0.2	5.2
Removal of Prostheses	3	0.1	1.3	122	0.0	0.5
Total Femoral	1	0.0	0.4	11	0.0	0.0
Reinsertion of Components	-	-	-	10	0.0	0.0
N Major	**90**	**2.5**	**40.0**	**10,990**	**1.7**	**45.5**
Insert Only	102	2.8	45.3	5820	0.9	24.1
Patella Only	21	0.6	9.3	4783	0.8	19.8
Insert/Patella	8	0.2	3.6	2479	0.4	10.3
Minor Components	3	0.1	1.3	48	0.0	0.2
Cement Only	1	0.0	0.4	11	0.0	0.0
N Minor	**135**	**3.7**	**60.0**	**13,141**	**2.1**	**54.5**
N Revision	**225**	**6.2**	**100.0**	**24,131**	**3.8**	**100.0**
N Primary	**3,652**			**636,393**		

Note. % Primaries Revised: The proportional contribution as a percentage of all primary procedures.
% Revisions: The number of revisions as a percentage of the total number of revisions.

5.6.2 Revision Rates of Comparator Groups by Fixation

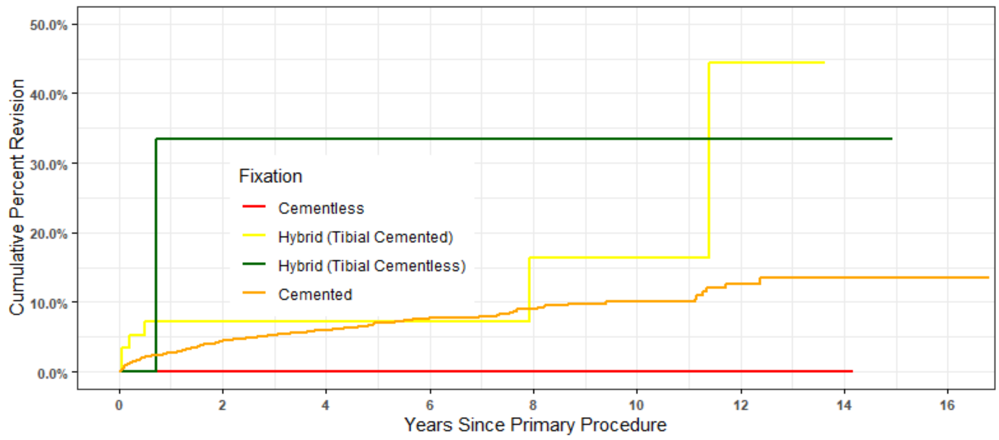
Prostheses typically have a recommended fixation method but can be used with an alternative fixation according to patient characteristics and the primary diagnosis. For primary conventional TKR, cementless fixation has a higher rate of revision than cemented fixation and hybrid (tibial-cemented) after three months. Figure 5.6 illustrates that hybrid (tibial-cementless) has the highest CPR up to 15.5 years of follow-up when a conventional stabilised knee is used. Significant differences were shown in the HR of the cementless against both cemented [HR 1.15 (1.12, 1.19), p<0.001] and hybrid (tibial-cemented) [HR 1.27 (1.23, 1.32), p<0.001] using Cox proportional hazard model. The cementless fixation with less than 10 observations has the lowest revision rate when a complex design prosthesis is used (Figure 5.7). However, this needs to be reassessed later over a larger sample size due to the limited numbers at risk. Cemented fixation shows better overall outcomes than both hybrid fixations for complex procedures. There is no significant difference in the HR between the revision rate of the fixation methods for complex procedures over the entire period.



Number at risk

Cementless (red)	119554	102974	87996	70818	52477	34483	20102	10129	2812
Hybrid (Tibial Cemented) (yellow)	146560	120129	94755	69898	48338	31666	19560	10069	2659
Hybrid (Tibial Cementless) (green)	6459	5815	4980	4093	3304	2386	1246	356	71
Cemented (orange)	363820	287674	217227	157723	109047	69022	38734	17579	4117

Figure 5. 6. CPR of conventional primary total knee replacement by fixation.



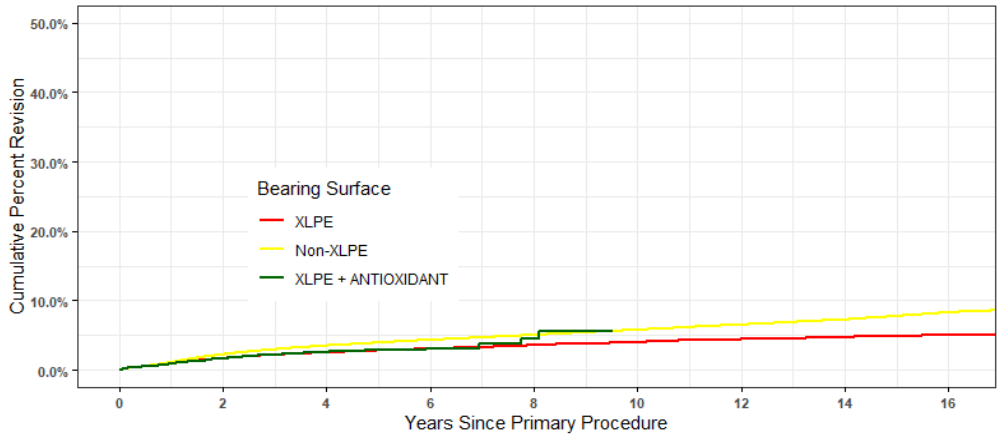
Number at risk

Cementless (red)	7	7	6	4	2	1	1	1	0
Hybrid (Tibial Cemented) (yellow)	57	38	24	17	9	4	2	0	0
Hybrid (Tibial Cementless) (green)	4	2	2	2	2	2	1	1	0
Cemented (orange)	3584	2463	1576	968	558	284	127	58	12

Figure 5. 7. CPR of complex primary total knee replacement by fixation.

5.6.3 Revision Rates of Comparator Groups by Bearing Surface

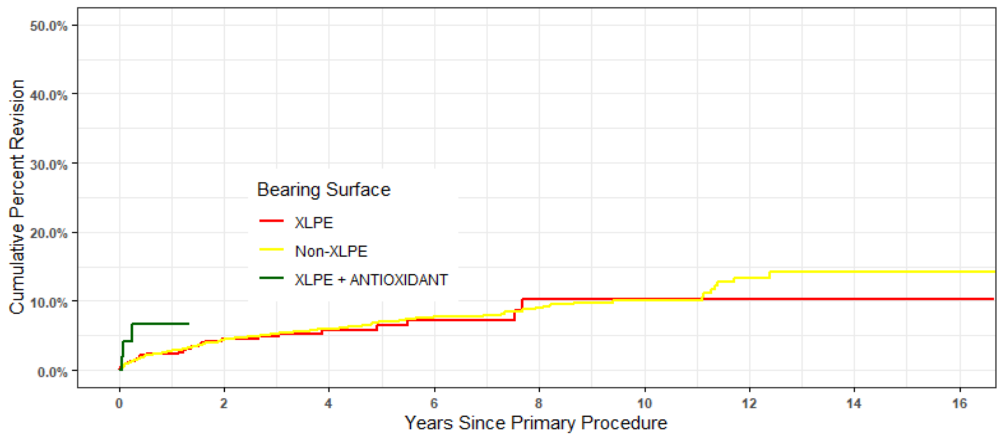
Two main polyethylene types are used in primary TKR: cross-linked polyethylene (XLPE) and non-XLPE. The XLPE includes a sub-group that has antioxidants added. After three months, XLPE has a lower revision rate than the non-XLPE for conventional procedures (Figure 5.8). The primary reason for this difference is a decreased cumulative incidence of late loosening. The difference between XLPE and non-XLPE is more apparent over a longer period. For complex procedures, there is a lower overall rate of revision for the XLPE compared to non-XLPE whereas the antioxidant version still needs more data to better reflect the revision outcome (Figure 5.9). The 15-year CPR rate of conventional procedures for XLPE is 4.92% (4.64, 5.21) and for non-XLPE is 7.66% (7.51, 7.81). Within the community of complex prostheses, the 15-year CPR for XLPE is 10.18% (4.97, 15.10) and for non-XLPE is 14.15% (10.67, 17.50). There are significant differences when non-XLPE is used compared to XLPE subtypes (including the antioxidant version) for the standard design prostheses [HR 1.40 (1.36, 1.44), p<0.001]. However, there was no statistical difference when XLPE was analysed with/without the addition of antioxidants. The same analysis for the complex procedures also shows no significant difference among the HRs of the types of bearing surfaces.



Number at risk

XLPE	230292	172586	118538	73335	40253	18877	7341	3408	741
Non-XLPE	344444	305802	262042	213279	161504	111243	68125	32892	8460
XLPE + ANTIOXIDANT	37206	16371	5411	454	97	0	0	0	0

Figure 5. 8. CPR of conventional primary total knee replacement by tibial bearing surface.



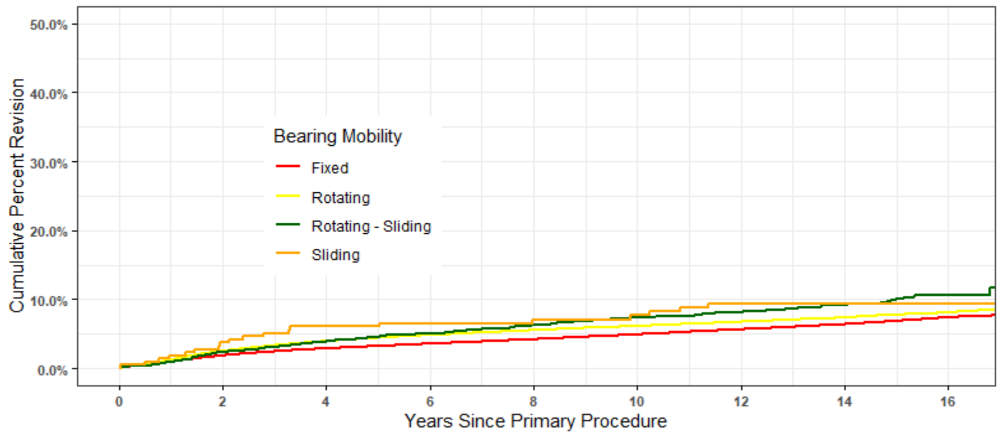
Number at risk

XLPE	489	318	178	102	51	27	11	5	3
Non-XLPE	3082	2179	1425	887	518	263	120	55	9
XLPE + ANTIOXIDANT	49	0	0	0	0	0	0	0	0

Figure 5. 9. CPR of complex primary total knee replacement by tibial bearing surface.

5.6.4 Revision Rates of Comparator Groups by Bearing Mobility

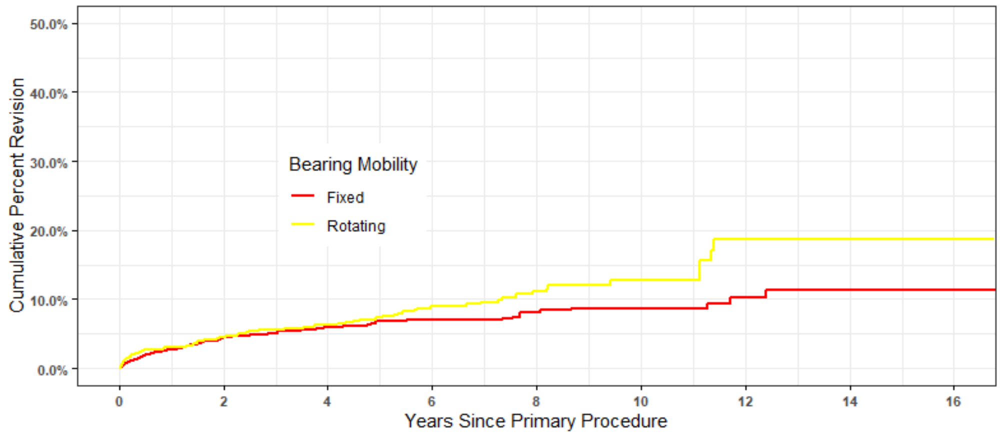
The CPR rate of complex and conventional TKR by bearing mobility is shown in Figures 5.10 and 5.11. Surgeons selected a variety of mobility designs with respect to patient characteristics. Tibial prostheses may be modular or non-modular. Modular prostheses with fixed or mobile designs have a metal base plate and tibial insert. Non-modular prostheses are either all-polyethylene or polyethylene moulded to a metal baseplate. Fixed bearings comprise non-modular tibial prostheses and those with fixed inserts that do not relatively move to the baseplate. Fixed-bearing prostheses have a lower overall CPR than all types of mobile bearings for complex and conventional procedures. Note that there is no complex primary TKR performed with rotating-sliding and sliding bearing mobility. When types of mobile bearings are compared for conventional prostheses, rotating mobility has a lower overall revision rate than the other types. However, the group of prostheses with sliding mobility design has been registered in only a limited number of observations. In total, there is a significant difference when comparing the combined group of mobile against fixed conventional prostheses using Cox proportional hazard model [HR 1.25 (1.21, 1.29), p<0.001].



Number at risk

512956	411136	317446	232649	159899	100171	56639	26846	6680
118494	101109	83744	66447	50165	34676	20749	9992	2502
4725	4140	3571	3248	2929	2547	2106	1168	431
218	207	197	188	173	163	148	127	46

Figure 5. 10. CPR of conventional primary TKR by bearing mobility.



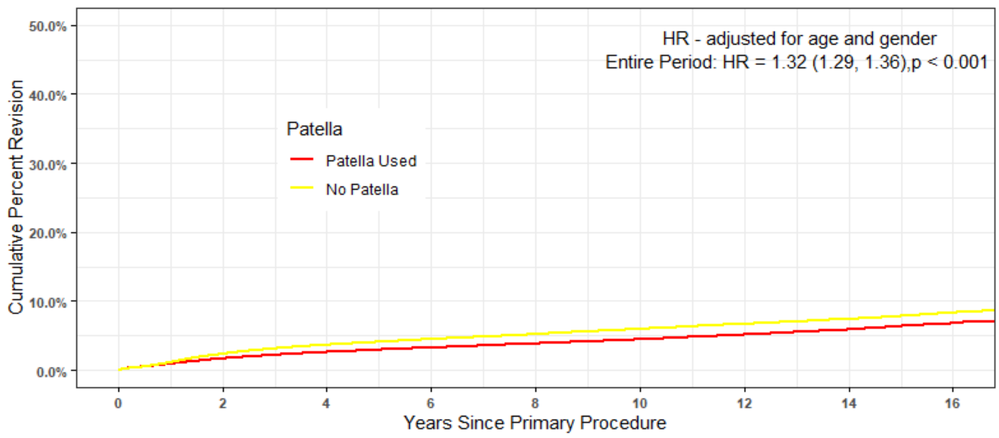
Number at risk

Fixed (red)	2475	1723	1095	648	367	197	93	43	8
Rotating (yellow)	1177	787	513	343	204	94	38	17	4

Figure 5. 11. CPR of complex primary TKR by bearing mobility.
(*Note*. There is no complex primary TKR performed with “Rotating - Sliding” and “Sliding” bearing mobility.)

5.6.5 Revision Rates of Comparator Groups by Patella Usage

Primary conventional TKR with patellar resurfacing has a lower revision rate significantly than procedures without patellar resurfacing [HR 1.32 (1.29, 1.36), p<0.001]. However, HR does not show a significant statistical difference when the patella is resurfaced for complex prostheses [HR 1.04 (0.79, 1.37), p=.78]. It is noted that outcomes related to the use of patellar resurfacing differ by the type of prosthesis used (Figures 5.12 and 5.13).



Number at risk

Patella Used (red)	358612	278199	207896	147986	99671	61799	35135	16584	4298
No Patella (yellow)	277781	238393	197062	154546	113495	75758	44507	21549	5361

Figure 5. 12. CPR of conventional primary TKR by patella usage.

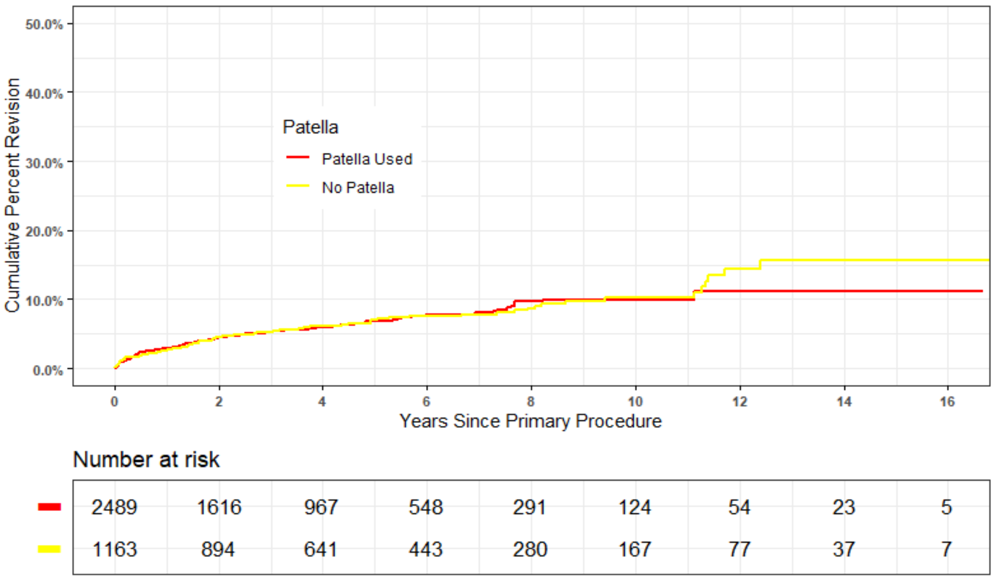


Figure 5. 13. CPR of complex primary TKR by patella usage.

5.7 Discussion

The problem with the current approach is that it does not adequately differentiate between the complex and conventional design prostheses. Given the higher revision risk of complex knee designs in primary total knee surgeries, this study aims to identify more relevant comparator groups to better reflect conventional and complex surgical practices. Conventional designs include CR and PS, and the knee designs used in more complex surgery includes fully stabilised (FS) and hinged designs.

When the CR and PS groups were combined as the final conventional comparator group, the 10-year CPR was 5.2% (5.1, 5.3). When the FS and hinged design groups were combined as a comparator group of complex devices to reflect devices used only for specific purposes, the CPR at 10-year was 10.3% (8.6, 12.0). The use of modified comparator groups led to identifying additional conventional prostheses but fewer complex designs as being at risk.

The conventional comparator improved sensitivity for the comparative assessment of standard design prostheses. In addition, a focus on complex prostheses generated a more relevant approach for the early identification of prostheses used for specific purposes in primary TKR. Through undertaking the AOANJRR standardised approach, fewer complex designs and additional conventional prostheses were identified as being at risk by utilising the modified comparator groups.

These findings may enhance the signal detection of poor-performing prostheses in a more relevant and effective comparative statistical analysis. Improved survivorship and better functional performance are projected when a new knee system surpasses a former model. However, these novel systems have design justifications to address stability, wear, and patellofemoral articulation. All design modifications do not deliver a consequence in improved survivorship [115]. Due to these ongoing changes to reduce complications, extend implant lifespan, and improve functional outcomes, the comparator needed to be reconsidered by the stability design to improve the relevance of comparative analyses.

This study also has several limitations. First, there was no further subdivision by other potential factors such as patella usage, fixation, bearing surface, and bearing mobility. However, each factor may influence the survivorship of comparator groups for complex and conventional designs [5, 34, 40]. At this stage, further subdivisions may adversely affect the effectiveness of initial screening for a conservative meaningful comparison of the prosthesis performance. Second, the AOANJRR has recently expanded classification to include medial pivot designs separately in its annual reports. This conventional design of total knee prostheses was not included in our scope as registered much less than CR and PS. The AOANJRR annual reports show that the medial pivot design provided satisfactory pain relief and functional improvement [5].

Some pre-existing conditions may affect the outcome of TKR because of the complexity of the surgery or the specific state of the affected limb. For example, when the surgeon is dealing with an unusual deformity of the bone or a soft-tissue envelope around the knee, there is the likelihood of increased risk of revision, thereby affecting the performance of individual prostheses [192-194]. Complex knees related to the type of patient or local conditions of the knee is considered and assessed by senior clinicians during the final stage of the AOANJRR standardised approach and are outside the scope of initial screening. Therefore, the focus of this chapter was on the first and second stages of the AOANJRR screening process. Future studies could be conducted to study a variety of factors, including the use of prostheses in complex primary situations, inadequate sample size, or whether they have been combined with prostheses already known to have a higher rate of revision or major differences in primary diagnosis.

The safety and effectiveness of medical devices such as knee arthroplasty prostheses are significant public health concerns [169]. Outlier detection will continue to evolve by reconsidering the improvements made periodically in prosthesis design and use. Joint registries play a significant role in controlling the harm and cost of using poor-performing devices in knee replacement surgeries [152]. An international collaboration between joint arthroplasty registries may enhance the process by generating a more comprehensive comparator for total complex and conventional knee prostheses [167].

5.8 Summary

This research suggests more relevant and effective comparator groups in primary TKR for a more appropriate comparison of device components. Utilising the conventional comparator improved the sensitivity for the comparative assessment of standard design prostheses. In addition, a focus on only complex prostheses generated a more specific approach for the early identification of prostheses used for specific purposes. The use of modified comparator groups led to identifying fewer complex and additional conventional prostheses as being at risk.

Chapter 6. Can Machine Learning Approach Contribute to Monitoring Post-Market Surveillance of Total Knee Arthroplasty Prostheses?

6.1 Overview

The industry continues to develop new implants and associated technologies, although more rigorous data is still needed to justify their introduction [165, 168, 169]. Total knee prostheses used in primary procedures are among the most relevant due to their widespread use and the number of poorly-performing devices [4, 190]. Monitoring the prostheses that have a higher risk of requiring revision, and early detection of these devices will produce better results in longer times and reduce health expenses [4, 152].

Most medical devices and surgical implants, including knee replacements, do not cause a problem or concern. However, joint replacement registries have played an important role in identifying the devices with a higher-than-anticipated revision rate called outliers [34, 44], particularly since it is difficult to ascertain the safety and comparative advantages of innovative knee implants that have been recently introduced into the market [152]. Hence, large-scale device evaluation using multinational registry data has become an essential means of determining whether a device itself has an increased risk of failure [167, 172].

A practical multistage approach has been developed by the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) to report the relative performance of prostheses [5]. Total knee devices are comprised of multiple parts working together, including two major femoral and tibial components. Revision surgery may occur because of the failure of one or both of these components. Both the tibial and femoral components used for primary total knee surgery are usually manufactured by the same company. This means that, generally, prosthesis combinations are identified by the AOANJRR rather than individual devices. The detection of total knee outliers with higher-than-anticipated revision rates is challenging as many prosthesis combinations are used differently depending on their purpose.

The aim of the initial screening method developed by the registry is to identify prostheses that differ significantly (twice than other total knee prostheses) from the combined revision rate per 100 observed component years for all other prostheses within the same class—comparator. The impact of confounding factors is also investigated by calculating the age- and gender-adjusted hazard ratio (HR) to determine whether there is a significant difference compared to the HR of the comparator. However, the method ignores the time ordering and does not address the confounding that may be due to device- and patient-related variables. Ideally, a time-to-event analysis needs to be undertaken to detect poor-performing devices while limiting the confounding effect of other components, device attributes, and patient characteristics.

Machine learning (ML) techniques are adopted for a variable selection problem because they have shown the potential to handle high-dimensional data with a vast number of interactions. This feature can be a solution to the additional complexity of confounding in medical research. The primary purpose of this thesis was to evaluate the use of ML for monitoring total knee prostheses. The effectiveness of the two proposed methods is determined by their ability to identify the same outliers identified by the AOANJRR standardised approach. The outcome can be used as a step towards improving post-market surveillance–evaluation efforts using AOANJRR-registry data.

6.2 Materials and Methods

The scope of this study is primary total knee replacement with primary diagnoses of Osteoarthritis (OA). The AOANJRR dataset contains 265,655 observations from the 1st of January 2015 (when the registry began collecting body mass index (BMI)) to the end of December 2019, as there was a desire to include all the possible patient-related confounding factors. Because almost all patients who underwent primary total knee arthroplasty had a major diagnosis of OA (98%), the data were filtered to only include cases with primary diagnosis OA. This comprehensive database comprises 160 unique prosthesis combinations, each of which consists of two major tibial and femoral components [5]. The same company's tibial/femoral components are typically combined and used in a total knee intervention, which was the reason for studying combinations rather than individuals [5].

Tibial/femoral components, device attributes, and patient characteristics are the inputs, and the time to the first revision surgery is the outcome of survival analysis. Each device component was clearly introduced to the model by an indicator variable showing the name of the model. Device attributes include prosthesis stability, bearing mobility, bearing surface, fixation, and use of patella (Table 6.1). The reason for using the covariates is to enable the algorithm to consider the linear and non-linear correlations. For example, the effect of fixation varies depending on prosthesis stability that needs to be considered in training [5].

Most patients were female (55.3%), had an American Society of Anesthesiologists (ASA) score of less than 3 (60.1%), the average age of 68.2 with a BMI of 32.1 kg/m². Patient covariates were age, gender, BMI, and American Society of Anesthesia (ASA) scores; all considered potential factors contributing to survival outcomes. Gender and ASA score (less than 3 vs. 3) were patient covariates defined in two levels. Age and BMI were categorized into three groups according to the distribution of our data. There were only missing values for BMI (6.26%) and ASA scores (0.41%) of patient data substituted using multiple imputation [174].

Table 6. 1. Descriptive information of patient- and device-related covariates.

Patient characteristics	Level	n (%)
Age	< 65	88,961 (33.5%)
	65-74	110,138 (41.5%)
	≥ 75	66,556 (25.0%)
Gender	Female	146,841 (55.3%)
	Male	118,814 (44.7%)
BMI	< 25	25,992 (9.8%)
	25-29.9	77,326 (29.1%)
	≥ 30	145,704 (54.8%)
ASA score	≥ 3	104,978 (39.5%)
	< 3	159,586 (60.1%)
Device attributes	**Level**	**n (%)**
Stability	Fully Stabilised	1,378 (0.5%)
	Hinged	628 (0.2%)
	Medial Pivot Design	19,918 (7.5%)
	Minimally Stabilised	183,470 (69.1%)
	Posterior Stabilised	60,261 (22.7%)
Bearing mobility	Fixed	230,106 (86.6%)
	Mobile	35,549 (13.4%)
Bearing surface	Non XLPE	103,947 (39.1%)
	XLPE	161,708 (60.9%)
Fixation	Cemented	177,522 (66.8%)
	Cementless	31,398 (11.8%)
	Hybrid	56,735 (21.4%)
Patella used	Yes	177,660 (66.9%)
	No	87,995 (33.1%)

Death was treated as a censored case, and survival time was calculated from the time of the primary implantation to the end of December 2019 for those patients who had not experienced revision or who had died. The main objective was to evaluate the use of ML to monitor primary total knee prostheses using data from the AOANJRR to control and reduce confounding. The exploration of the variable importance using ML techniques needs a model that includes carefully-selected hyperparameters [195]. A two-step ML approach was proposed for both the random survival and Cox models to ascertain the significance of variables in the presence of confounding effects. The effectiveness was determined by the model's ability to detect the same outliers identified after *I* and *II* stages of the AOANJRR standardised approach.

The first stage of the standard is a screening test that identifies prostheses that exceed twice the combined revision rate per 100 observed component years of all other prostheses in the same class. In the second stage, age- and gender-adjusted HR is calculated to check whether there is a significant difference compared to the combined hazard rate of the comparator and take into consideration the impact of confounding. The comparator for the conventional practices contains all other prostheses designed with only PS and CR stability. In addition, the comparator for complex procedures used only for specific purposes in primary total knee replacement involves FS and hinged designs. These comparator groups are the modified final versions developed in Chapter 5.

R Statistical software glmnet package [185] version 4.1-1 was used for the Cox elastic net, survival package [186] version 3.2-11 for the unregularized Cox and randomForestSRC [174] version 2.11.0 for random survival modelling and MICE package [187] version 3.14.0 for multiple imputations of missing values.

6.3 Machine Learning Statistical Analyses

The first method used was an extension of random forest called Random Survival Forest (RSF) to analyse survival data with right-censored cases [175, 176]. A forest is a group of 2,000 trees; each tree grows by iterating the binary split of the AOANJRR data using a log-rank test until a stopping rule is reached. A random set of variables splits the candidate-selected node into two daughter nodes from each parent node. This variable maximizes the log-rank statistic [177] until a terminal node has no fewer than two revisions. Because of a focus on feature selection rather than prediction, this study chose deep trees to improve the probability of reflecting variations between the predictors.

Variable selection is randomized using the parameter *'mtry'*. The algorithm selects *X* of the maximum of the input variables (*P*) randomly on each node. The variables considered in each division of each tree were randomly selected, but the number of variables was fixed at ' $P/4$ ' [116]. The number of variables considered in each splitting is greater than the conventional (√*P*), as an increasing number of variables in each split is able to limit the bias in the selection of correlated input variables [178].

The method excluded noisy variables using a backward selection procedure in order to determine the most important variables. To obtain a reduced set of salient variables, the following stepwise selection method was implemented to systematically remove noise variables: (i) calculate RSF using all the covariates and all device components; (ii) evaluate the inputs and remove noisy variables; (iii) compute a new RSF using the remaining variables; and (iv) select a set of components with a higher risk of revision. Finally, a *P*-value cut-off of 0.05 was selected to characterize outliers. A similar algorithm was proposed by Ishwaran et al. [121] and Dietrich et al. [179].

Variables were ranked according to the minimal depth [180]. The minimal depth of a variable is the distance from the tree's root node to the node where the variable is split first. The distance is recorded for each variable, and then their average within the forest is computed. Shorter distances show variables with more significant impacts. To determine whether the minimum depth of the device component exceeds the probability, a threshold *P*-value of 0.05 was determined according to the empirical null distribution for each variable [116, 181]. The null distribution is based on 1,000 permutations of the response, grows a forest with 200 trees in each, and calculates the minimal depth of each variable. The adjusted *P*-values based on false discovery rate (FDR) were not calculated because of the small number of permutations implemented while it would incur a higher computational cost. A variable is considered significant if the permutation *P*-value was less than 0.05.

Secondly, regularized/unregularized Cox was applied using an ML supervised algorithm combined with a well-known conventional method. The second step suggests a more understandable approach for the interpretation of outcomes and the reporting of statistical significance of inputs. Some of the device components that best predicted survival were selected using a standardized model with a combination of L1 (lasso) and L2 (ridge) penalties. The elastic-net in the presence of lasso or ridge was chosen due to its superior performance with highly-correlated variables [182, 183]. The elastic-net was specified by a value ($\alpha = 0.5$; ranged from 0 to 1) between ridge regression ($\alpha = 0$) and LASSO ($\alpha = 1$). The parameters determining the complexity of the model were chosen by 10-fold Cross-Validation [182]. No penalty was applied to any of the four patient variables in the model, as the intention was to fully control the effects of a relatively small number of patient characteristics.

The regularized Cox model does not report a *P*-value because it does not test the variable for the null hypothesis. This was the reason for using the second step, where the selected variables are included in the unregularized Cox proportional hazards model. Given a need to draw inferences while appreciating that a selection process was initially undertaken, *P*-values that maintain the FDR at 0.05 [184] were also discovered. This was done by using the total number of device components that were included in the regularized model (we set *P*-values ≈ 1 for unselected variables, as implied by zero coefficient in the model). Control over the FDR keeps the portion of false discoveries at the chosen nominal value among the rejected null hypotheses.

6.4 Results

Figure 6.1 shows the survival of devices over the period chosen for this study. Initially, the AOANJRR standard was employed as the metric to evaluate the performance of ML in the initial screening of knee prostheses. The prostheses identified as having higher-than-expected rates of revision according to the AOANJRR standard are listed in Table 6.2. The AOANJRR standardised approach identified five conventional/non-complex design combinations. From the prostheses listed below, the registry has reported all these devices through previous annual investigations with a greater number of observations. A device with a complex design is not generally identified as an outlier after all the stages due to an expected higher risk of revision for the use of these devices in primary total knee surgery.

Table 6. 2. Identified outlier combinations by the 1st and 2nd stages of the AOANJRR standard.

Femoral/Tibial	**Descriptive information**			**1st stage**	**2nd stage**	**Comparator**
Conventional	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs.Years (95% CI)**	**HR - adjusted for age and gender, *P*-value**	**Other total knee (PS & CR)**
Device *I*	19	401	1068.8	1.78 (1.07, 2.78)	2.33 (1.48, 3.65) p<0.001	0.76 (0.74, 0.78)
Device *II*	25	561	1405.8	1.78 (1.15, 2.62)	2.23 (1.51, 3.31) p<0.001	0.76 (0.74, 0.78)
Device *III*	22	436	1416.4	1.55 (0.97, 2.35)	2.07 (1.36, 3.15) p<0.001	0.76 (0.74, 0.78)
Device *IV*	116	2648	7109.6	1.63 (1.35, 1.96)	2.21 (1.84, 2.66) p<0.001	0.75 (0.73, 0.77)
Device *V*	15	222	560.1	2.68 (1.50, 4.42)	3.28 (1.97, 5.44) p<0.001	0.76 (0.74, 0.78)

Note. The comparator for conventional prostheses includes all other prostheses with PS and CR stability, and the comparator used for complex prostheses involves only FS and hinged designs.

Table 6.3 indicates the extent to which outliers were detected using the RSF backward selection procedure and regularized/unregularized Cox. Devices *IV* and *V* were identified using both the ML methods, taking into account patient- and device-related confounding. However, only one of the same outliers (*IV*) was detected when the FDR was maintained at 0.05 by regularized Cox. The Cox approach showed greater performance by reporting an additional device (*III*).

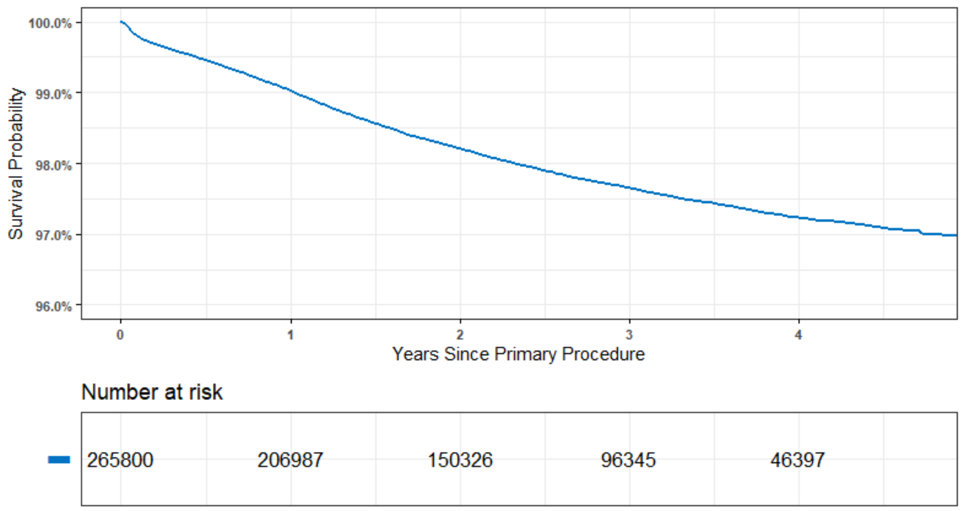


Figure 6. 1. Time to first revision for the AOANJRR primary total knee dataset.

In the case of RSF, closer proximity to the root node means a smaller average minimal depth and indicates a more significant contribution by the predictor. However, a lower minimal depth rank does not necessarily imply a poorer prosthesis. Since RSF cannot report the exact *P*-value, the distribution of importance under the null hypothesis of no association to the response was created by several replications of permutation responses. Note that the noisy variables were removed in an initial step and were not included in the second RSF model in order to avoid bias.

Table 6. 3. Results for identified prosthesis combinations by RSF and regularised/unregularised Cox.

Femoral/Tibial	**Descriptive information**			**Random Survival Forest**	**Regularized/Unregularized Cox**
Conventional	**N Revised**	**N Total**	**Obs.Years**	**Minimal depth rank *P*-value**	***P*-value**
Device *I*	19	401	1068.8	-	-
Device *II*	25	561	1405.8	-	-
Device *III*	22	437	1417.8	-	0.018
Device *IV*	116	2648	7109.6	4 P=0.009	p<0.001
Device *V*	15	222	560.1	6 P=0.009	0.004

Note. Regularized Cox model selected 85 components. *P*-values reported for the Cox technique are based on a Wald test from an unregularized Cox model with the selected variables. The rank column is based on the values for minimal depth. Ranks closer to zero indicate smaller minimal depths representing more substantial variable effects. In the case of the regularized/unregularized Cox, “-” signifies that no *P*-value is provided, and the prosthesis was not selected by the model. For the RSF, "-" means that the prosthesis was not included in any trees of the second forest after removing noisy variables; thus, there is no rank or *P*-value.

The AOANJRR gold standard provided an ideal means of evaluating the performance of ML in detecting outlier prostheses. Although the Cox method identified three of the same prostheses, only one of them was detected after controlling for the FDR. Due to the data-dependent nature of ML, a limited number of observations imposes practical constraints on the identification. The main reason for using ML is to control potential confounding. This was evaluated by comparing HR for specific components in two models: (*a*) the age and gender-adjusted Cox proportional hazard model with a variable representing the use of that component; and (*b*) the unregularized Cox model that includes all the variables selected using the elastic-net (i.e., when it was conditioned on the other components and selected covariates). Therefore, the difference in HRs between these two models indicates the extent of potential confounding with respect to the AOANJRR standard. Most prostheses have some reasonable evidence of confounding. The relative difference of 34%, 9%, and 43% in model coefficients are shown in Figure 6.2 for Devices *III*, *IV*, and *V,* respectively.

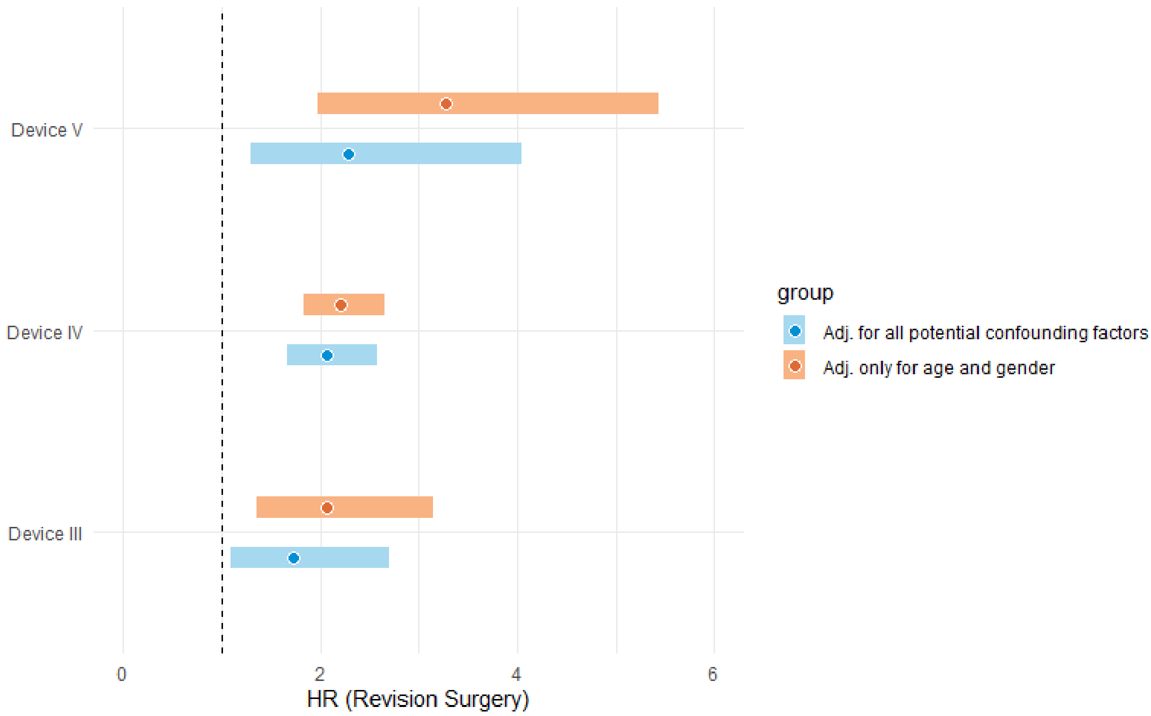


Figure 6. 2. HR comparison to illustrate the potential confounding.
(*Note.* %Diff = [ln ($HR_{\text{Adj. for age and gender}}$) – ln ($HR_{\text{Adj. for all potential confounding factors}}$)]/[ln ($HR_{\text{Adj. for all potential confounding factors}}$)])

An additional step (stage *III*) of the AOANJRR standard allows clinicians to obtain further information about the identified prosthesis derived from stages *I* and *II*. A full range analysis is conducted, and the results are discussed during a two-day workshop before the outliers are documented in the annual reports. Workshop participants consider the factors that could contribute to potential errors, such as the effect of a single surgeon, range of catalogue numbers, or a device having a specific purpose. This stage enables more real outliers to be recognised and additional confounding factors to be considered, supported by more robust evidence. As a result, this study designed a compelling and particular method for the comparative study of newly-introduced prostheses with the view to assessing the impact of surgeons and catalogue ranges as a post-statistical analysis. The use of a current comparator leads to the identification of much more complex prostheses during the first and second stages, although most of them might not be true outliers as they are mostly used for specific purposes with an expected higher rate of revision.

6.5 Subsets of Prosthesis – Catalogue Ranges

There is a solid argument for using the identification method to examine catalogue ranges so as to detect the type of prosthesis that is likely to have a higher revision rate than the comparator. The results presented in Tables 6.5 and 6.7 show that Devices *I* and *II* might not be poor-performing prostheses, as only a subset of these prostheses has issues. There are significant variations in the performance of prosthesis subset designs (Tables 6.4-6.13). This shows strong evidence to suggest the role of confounding factors in detecting poor-performing prostheses. The clinicians should be informed about the current status of prostheses using a further analysis of all the potential confounders.

Table 6. 4. Catalogue range description of Device *I* primary total knee replacement.

Catalogue Range	Catalogue Description	Cement	Coating	Fixation
Femoral				
KFTCPC0L-KFTCPC6R	CR Porous Primary Femoral Component (Wright)	NO	-	POROUS
KFTCHA1L-KFTCHA6R	CR HA Primary Femoral Component (Wright)	NO	HA Coated	POROUS
KFTCPN2L-KFTCPN4R	Advance Stature CR Porous Femoral Component (Wright)	NO	-	POROUS
KFTCPC0L-KFTCPC6R	CR Porous Primary Femoral Component (MicroPort)	NO	-	POROUS
KFTCPN2L-KFTCPN4R	Advance Stature CR Porous Femoral Component (MicroPort)	NO	-	POROUS
Tibial				
KTSCFM10-KTSCFM51	BioFoam Tibial Base w/ Screwholes Ti (Wright)	NO	-	-
KTSCFM10-KTSCFM51	BioFoam Tibial Base w/ Screwholes Ti (MicroPort)	NO	-	-

Table 6. 5. Revision rates of Device *I* by catalogue number range.

Femoral range	Tibial range	N Revised	N Total	Obs.Years	Revisions/100 Obs.Yrs (95% CI)
KFTCPC0L-KFTCPC6R	KTSCFM10-KTSCFM51	7	89	372.8	1.88 (0.75, 3.87)
KFTCPC0L-KFTCPC6R	KTSCFM10-KTSCFM51	0	3	14.0	0.00 (0.00, 26.37)
KFTCPN2L-KFTCPN4R	KTSCFM10-KTSCFM51	0	2	5.0	0.00 (0.00, 74.22)
KFTCPC0L-KFTCPC6R	KTSCFM10-KTSCFM51	1	16	58.3	1.71 (0.04, 9.56)
KFTCPC0L-KFTCPC6R	KTSCFM10-KTSCFM51	0	16	50.9	0.00 (0.00, 7.24)
KFTCPN2L-KFTCPN4R	KTSCFM10-KTSCFM51	0	5	15.7	0.00 (0.00, 23.44)
KFTCPC0L-KFTCPC6R	KTSCFM10-KTSCFM51	11	259	535.7	2.05 (1.02, 3.87)
KFTCPN2L-KFTCPN4R	KTSCFM10-KTSCFM51	0	11	16.3	0.00 (0.00, 22.63)
Total		**19**	**401**	**1068.8**	**1.78 (1.16, 2.61)**

Table 6. 6. Catalogue range description of Device *II* primary total knee replacement.

Catalogue Range	Catalogue Description	Cement	Coating
Femoral			
NO582K-NO688K	FP/UC Cementless Femoral Component	NO	HA COATED
NB702K-NB758K	PS CoCr Cemented Standard Femoral Component	YES	-
NO502Z-NO608Z	AS FP/UC Cemented Premium Femoral Component	YES	-
NB702Z-NB758Z	AS PS CoCr Cemented Premium Femoral Component	YES	-
NO502K-NO608K	FP/UC Cemented Standard Femoral Component	YES	-
Tibial			
NB731Z-NB788Z	AS UC/PS Cemented Modular Tibial Plateau	YES	-
NB741K-NB798K	UC/PS Cementless Modular Tibial Plateau	NO	HA COATED
NX731K-NX788K	UC/PS Cemented Pro Modular Tibial Component	YES	-
NB731K-NB788K	UC/PS Cemented Modular Tibial Plateau	YES	-

Table 6. 7. Revision rates of Device *II* by catalogue number range.

Femoral range	Tibial range	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
NO582K-NO688K	NB731Z-NB788Z	0	2	6.9	0.00 (0.00, 53.08)
NO582K-NO688K	NB741K-NB798K	14	130	458.3	3.05 (1.67, 5.12)
NO582K-NO688K	NX731K-NX788K	0	56	85.7	0.00 (0.00, 4.30)
NO582K-NO688K	NB731K-NB788K	0	1	4.8	0.00 (0.00, 76.37)
NB702K-NB758K	NB731Z-NB788Z	2	94	151.6	1.32 (0.16, 4.76)
NB702K-NB758K	NX731K-NX788K	0	15	4.0	0.00 (0.00, 92.68)
NO502Z-NO608Z	NB731Z-NB788Z	3	159	407.0	0.74 (0.15, 2.15)
NO502Z-NO608Z	NB741K-NB798K	0	1	4.2	0.00 (0.00, 88.67)
NB702Z-NB758Z	NB731Z-NB788Z	0	1	2.3	0.00 (0.0, 161.79)
NB702Z-NB758Z	NX731K-NX788K	0	4	19.3	0.00 (0.00, 19.11)
NO502K-NO608K	NX731K-NX788K	6	98	261.7	2.29 (0.84, 4.99)
Total		**25**	**561**	**1405.8**	**1.78 (1.24, 2.48)**

Table 6. 8. Catalogue range description of Device *III* primary total knee replacement.

Catalogue Range	Catalogue Description	Cement
Femoral		
196008400-196009400	PS RPF CoCr Femoral Component	YES
196040100-196050600	PS Cemented Femoral Component	YES
196004400-196005400	CS Cemented Femoral Component	YES
196081100-196091600	PS150 CoCr High Flex Femoral Component	YES
960042-960058	Cruciate Sacrificing NonPorous Femoral Component	YES
950010-950027	RPF CoCr Cemented Femoral Component	YES
Tibial		
129433110-129433170	Cemented Keel Tibial Tray	YES
129435215-129435415	Revision Cemented 15mm Tibial Tray	YES
129431110-129431170	Cemented Tibial Tray	YES
129435110-129435160	CoCr Revision Cemented Tibial Tray	YES

Table 6. 9. Revision rates of Device *III* by catalogue number range.

Femoral range	Tibial range	N Revised	N Total	Obs.Years	Revisions/100 Obs.Yrs (95% CI)
196008400-196009400	129433110-129433170	1	21	79.4	1.26 (0.3, 9.10)
196040100-196050600	129433110-129433170	4	46	119.7	3.34 (0.91, 8.56)
196040100-196050600	129435215-129435415	0	2	5.7	0.00 (0.00, 65.06)
196040100-196050600	129431110-129431170	4	50	94.4	4.24 (1.15, 10.85)
196040100-196050600	129431110-129431170	0	1	3.4	0.00 (0.0, 107.23)
196040100-196050600	129435110-129435160	4	77	194.2	2.06 (0.56, 5.27)
196004400-196005400	129431110-129431170	1	3	11.3	8.88 (0.22, 49.48)
196081100-196091600	129433110-129433170	0	16	56.4	0.00 (0.00, 6.54)
196081100-196091600	129431110-129431170	0	1	1	0.00 (0.0, 380.3)
960042-960058	129431110-129431170	2	133	526.9	0.38 (0.04, 1.37)
960042-960058	129435110-129435160	2	7	26.4	7.56 (0.91, 27.31)
950010-950027	129433110-129433170	4	80	299	1.34 (0.36, 3.42)
Total		**22**	**436**	**1416.4**	**1.55 (0.97, 2.35)**

Table 6. 10. Catalogue range description of Device *IV* primary total knee replacement.

Catalogue Range	Catalogue Description	Cement	Coating
Femoral			
10200201-10200217	CoCr Min. Stab. Femoral Component	YES	-
10200101-10200117	CoCr Min Stab. HA Pegged Stippled Surface Femoral Component	NO	HA COATED
Tibial			
10200501-10200507	CoCr Polished Tibial Baseplate	YES	-
10200401-10200407	CoCr HA Stippled Surface Tibial Baseplate	NO	HA COATED

Table 6. 11. Revision rates of Device *IV* by catalogue number range.

Femoral range	Tibial range	N Revised	N Total	Obs.Years	Revisions/100 Obs.Yrs (95% CI)
10200201-10200217	10200501-10200507	14	482	1323.2	1.06 (0.58, 1.77)
10200201-10200217	10200401-10200407	0	3	9.6	0.00 (0.00, 38.27)
10200101-10200117	10200501-10200507	51	1212	3338.5	1.53 (1.14, 2.01)
10200101-10200117	10200401-10200407	51	951	2438.2	2.09 (1.56, 2.75)
Total		**116**	**2648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 6. 12. Catalogue range description of Device *V* primary total knee replacement.

Catalogue Range	Catalogue Description	Cement	Fixation
Femoral			
184500-184536	Vanguard PS Open Box Femoral Porous Coated/Bond Coated	NO	POROUS
183100-183136	Vanguard PS Open Box Femoral Interlok	YES	MATT
Tibial			
141270-141278	Porous Tibial Tray	NO	POROUS

Table 6. 13. Revision rates of Device *V* by catalogue number range.

Femoral range	Tibial range	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
184500-184536	141270-141278	2	48	154.0	1.30 (0.16, 4.69)
183100-183136	141270-141278	13	174	406.1	3.20 (1.70, 5.47)
Total		**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

6.6 Number of Surgeons

The registry is aware that a single surgeon may be responsible for a prosthesis that has a higher revision rate. This situation has occurred twice, and on both occasions, subsequent use of the femoral/tibial combination ceased following the publication of the Annual Report [5, 165]. Revision rates per 100 component years of the identified prostheses were evaluated to determine the effect that a surgeon had on the performance of prostheses. By investigating the revision rate of surgeons for each identified device component, those who did not perform well with a device of interest can be determined (Table 6.14-6.18). The tables below show surgeon *ID*s and the results of surgical procedures.

This analysis shows the most significant interactions of devices and surgeons using the first stage of the AOANJRR standardised approach. Only the surgeons who had performed more than 10 procedures using a device of interest were included. This status is significant for two reasons: (*i*) some surgeons may show a relatively acceptable performance with the same device component, or (*ii*) in the case when a device is still not being used to a great extent by various surgeons. All these knee prostheses show significant interaction with the experience and expertise of the surgeon. For example, surgeons 1177, 1218, 1745 could be responsible for a higher than expected rate of revision for Device *IV,* or the device might affect the performance of the surgeon (Table 6.14).

Table 6. 14. Revision rates of Device *I* primary total knee replacement by surgeon id.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
283	Device *I*	2	134	335.1	0.59 (0.07, 2.16)
	Overall	3	163	407.1	0.74 (0.15, 2.15)
	Ex. Outliers	1	29	72	1.39 (0.03, 7.73)
482	Device *I*	17	264	723.0	2.35 (1.37, 3.76)
	Overall	18	281	790.6	2.28 (1.35, 3.6)
	Ex. Outliers	1	17	67.7	1.48 (0.04, 8.23)

Table 6. 15. Revision rates of Device *II* primary total knee replacement by surgeon id.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
101	Device *II*	0	11	53.1	0.00 (0.00, 6.95)
	Overall	4	437	1224.9	0.33 (0.09, 0.84)
	Ex. Outliers	4	426	1171.8	0.34 (0.09, 0.87)
153	Device *II*	16	247	652.7	2.55 (1.40, 3.98)
	Overall	17	281	751.5	2.26 (1.32, 3.62)
	Ex. Outliers	0	11	9.3	0.00 (0.00, 39.54)
975	Device *II*	0	48	166.5	0.00 (0.00, 2.21)
	Overall	8	297	690.3	1.16 (0.50, 2.28)
	Ex. Outliers	7	237	492.3	1.42 (0.57, 2.93)
1037	Device *II*	2	107	276	0.72 (0.09, 2.62)
	Overall	2	107	276	0.72 (0.09, 2.62)
	Ex. Outliers	-	-	-	-
1070	Device *II*	2	115	159.2	1.26 (0.15, 4.54)
	Overall	4	272	674	0.59 (0.16, 1.52)
	Ex. Outliers	2	157	514.9	0.39 (0.05, 1.4)
1290	Device *II*	1	15	54.7	1.82 (0.05, 10.18)
	Overall	2	72	111.8	1.79 (0.22, 6.46)
	Ex. Outliers	1	57	57.1	1.75 (0.04, 9.75)

Table 6. 16. Revision rates of Device *III* primary total knee replacement by surgeon id.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
3	Device *III*	2	19	63.0	3.17 (0.38, 11.46)
	Overall	169	7408	21454.9	0.79 (0.67, 0.91)
	Ex. outliers	166	7364	21319.3	0.78 (0.66, 0.90)
266	Device *III*	0	16	45.5	0.00 (0.00, 8.11)

	Overall	12	356	814.8	1.47 (0.76, 2.57)
	Ex. outliers	12	340	769.3	1.56 (0.80, 2.72)
431	Device *III*	3	80	239.8	1.25 (0.26, 3.65)
	Overall	3	80	239.8	1.25 (0.26, 3.65)
	Ex. outliers	-	-	-	-
651	Device *III*	1	14	33.5	2.98 (0.07, 16.63)
	Overall	7	603	1518.6	0.46 (0.18, 0.95)
	Ex. outliers	6	589	1485.1	0.40 (0.15, 0.88)
810	Device *III*	5	99	368.7	1.35 (0.44, 3.16)
	Overall	9	255	629.9	1.43 (0.65, 2.71)
	Ex. outliers	4	156	261.2	1.53 (0.42, 3.92)
1167	Device *III*	2	14	38.8	5.16 (0.62, 18.63)
	Overall	4	105	245.3	1.63 (0.44, 4.17)
	Ex. outliers	2	91	206.5	0.97 (0.12, 3.50)
1372	Device *III*	1	11	28.2	3.55 (0.09, 19.76)
	Overall	18	804	1622.1	1.11 (0.66, 1.75)
	Ex. outliers	17	793	1593.9	1.07 (0.62, 1.71)
1434	Device *III*	0	12	43.1	0.00 (0.00, 8.55)
	Overall	2	156	331	0.60 (0.07, 2.18)
	Ex. outliers	2	144	287.9	0.69 (0.08, 2.51)
1721	Device *III*	2	79	315.1	0.63 (0.08, 2.29)
	Overall	5	206	515	0.97 (0.31, 2.26)
	Ex. outliers	3	127	199.9	1.50 (0.31, 4.39)

Table 6. 17. Revision rates of Device *IV* primary total knee replacement by surgeon id.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
57	Device *IV*	1	28	125.5	0.8 (0.02, 4.44)
	Overall	9	630	1496.1	0.6 (0.27, 1.14)
	Ex. outliers	8	602	1370.6	0.58 (0.25, 1.15)
153	Device *IV*	1	23	89.4	1.12 (0.03, 6.23)
	Overall	17	281	751.5	2.26 (1.32, 3.62)
	Ex. outliers	0	11	9.3	0.00 (0.00, 39.54)
173	Device *IV*	1	45	146.3	0.68 (0.02, 3.81)
	Overall	10	352	864.3	1.16 (0.55, 2.13)
	Ex. outliers	9	307	717.9	1.25 (0.57, 2.38)
275	Device *IV*	4	143	516.5	0.77 (0.21, 1.98)
	Overall	4	218	610.9	0.65 (0.18, 1.68)
	Ex. outliers	0	75	94.5	0.00 (0.00, 3.90)
282	Device *IV*	6	152	618	0.97 (0.36, 2.11)
	Overall	11	787	1993.2	0.55 (0.27, 0.99)
	Ex. outliers	5	635	1375.2	0.36 (0.12, 0.85)
456	Device *IV*	0	10	12.3	0.00 (0.00, 30.09)
	Overall	3	132	374.1	0.80 (0.16, 2.34)
	Ex. outliers	3	122	361.8	0.83 (0.17, 2.42)
495	Device *IV*	9	191	623.9	1.44 (0.66, 2.74)
	Overall	12	369	926.9	1.29 (0.67, 2.26)
	Ex. outliers	3	178	303	0.99 (0.20, 2.89)
895	Device *IV*	3	85	332.9	0.9 (0.18, 2.63)
	Overall	9	773	1763.3	0.51 (0.23, 0.97)
	Ex. outliers	6	688	1430.4	0.42 (0.15, 0.91)
934	Device *IV*	0	11	34.3	0.00 (0.00, 10.75)
	Overall	12	644	1498.7	0.8 (0.41, 1.40)
	Ex. outliers	12	633	1464.4	0.82 (0.42, 1.43)
961	Device *IV*	5	159	243.6	2.05 (0.67, 4.79)
	Overall	5	202	422.3	1.18 (0.38, 2.76)
	Ex. outliers	0	43	178.7	0.00 (0.00, 2.06)
1009	Device *IV*	2	58	239.5	0.83 (0.10, 3.02)
	Overall	18	1116	2610.4	0.69 (0.41, 1.09)
	Ex. outliers	16	1058	2370.9	0.67 (0.38, 1.09)
1119	Device *IV*	34	511	1637.8	2.07 (1.44, 2.9)
	Overall	37	826	2046.3	1.81 (1.27, 2.49)
	Ex. outliers	3	315	408.4	0.73 (0.15, 2.15)
1149	Device *IV*	16	379	716.9	2.23 (1.27, 3.62)
	Overall	23	566	1390.9	1.65 (1.05, 2.48)
	Ex. outliers	7	187	674	1.04 (0.42, 2.14)
1177	Device *IV*	15	197	649.9	2.31 (1.29, 3.81)
	Overall	21	219	721.5	2.91 (1.80, 4.45)
	Ex. outliers	6	22	71.6	8.38 (3.07, 18.24)
1195	Device *IV*	0	38	111.6	0.00 (0.00, 3.30)

	Overall	17	1154	2811.5	0.60 (0.35, 0.97)
	Ex. outliers	17	1116	2700	0.62 (0.37, 1.01)
1218	Device *IV*	7	439	414.4	1.69 (0.68, 3.48)
	Overall	39	921	1963.8	1.98 (1.41, 2.71)
	Ex. outliers	32	482	1549.4	2.06 (1.41, 2.91)
1258	Device *IV*	0	17	64.6	0.00 (0.00, 5.71)
	Overall	1	48	98.6	1.01 (0.02, 5.65)
	Ex. outliers	1	31	33.9	2.94 (0.07, 16.41)
1745	Device *IV*	12	106	374.5	3.2 (1.65, 5.6)
	Overall	15	231	561.7	2.67 (1.49, 4.40)
	Ex. outliers	3	125	187.2	1.60 (0.33, 4.68)
1810	Device *IV*	0	20	72.1	0.00 (0.00, 5.11)
	Overall	6	219	745.6	0.80 (0.29, 1.75)
	Ex. outliers	6	199	673.4	0.89 (0.33, 1.94)

Table 6. 18. Revision rates of Device *V* primary total knee replacement by surgeon id.

Surgeon ID		N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
253	Device *V*	2	42	135.9	1.47 (0.18, 5.32)
	Overall	10	745	1742.5	0.57 (0.27, 1.05)
	Ex. outliers	8	703	1606.6	0.50 (0.21, 0.98)
660	Device *V*	13	172	401.2	3.24 (1.72, 5.54)
	Overall	17	260	670.2	2.54 (1.48, 4.06)
	Ex. outliers	4	88	269	1.49 (0.40, 3.81)

6.7 Discussion

The results of this chapter suggest that the proposed ML methods may be effective for the detection of poor-performing arthroplasty devices. ML methods do not require the analyst to know in advance the linear and non-linear relationships between variables over time. However, one single model could not be necessarily the best method to deal with a high-dimensional dataset with multicollinearity in a feature selection problem. Any inclusion or exclusion of noisy variables influences the outcome in terms of a given predictor's strength and sign. A backward selection step was added to illustrate the applicability of RSF to exploratory data analysis in a prospective cohort [179, 196]. This method allowed us to reduce the dimensions of a complex dataset and improve variable selection since many noisy variables are excluded.

In some cases, the prosthesis identified after the second stage of the AOANJRR standard may not be a real outlier due to a single surgeon, or it might be a poor-performing subset of a device with a specific attribute and design. Although the registry does not identify these prostheses, their ongoing real-time performance is monitored. There is evidence of the issue of having only a subset of unidentified prostheses (Device *I* and *II*) from the AOANJRR online documents provided to further indicate the performance of total knee prostheses. In addition, our investigation of the catalogue ranges showed significant variations in the revision rates with respect to design features.

Regularized/unregularized Cox produced results more comparable to the AOANJRR standard in detecting the same total knee devices, whereas there was not a significant difference in the performance of the proposed methods. Using an unregularized step for the Cox enabled us to apply a more conventional method of documenting confounding and reducing the mistrust in the black-box nature of ML analysis. To handle multicollinearity which may increase the risk of selecting an arbitrary predictor [196], the two-step approach was proposed for RSF modelling to reduce the dimension by removing the noisy variables. In addition, RSF grows the tree using a random subset of variables per node and splitting the nodes by independently selecting the input variables [188]. This feature provides random survival attractive for exploratory variable selection, where false-positive discovery due to overfitting is still an important issue [189].

One of the key aims of the current work was to limit the confounding that was handled by both the ML methods used. The regularized/unregularized Cox conditions simultaneously in device- and patient-related factors. With the RSF approach, it is assumed that many variables are competing at the same time in the splitting process. The Cox approach showed significant advantages in terms of reducing computational time, interpreting variable effects, and documenting confounding effects. The random survival reduces variance and bias using many variables and automatically assesses linear or complex non-linear interactions. Various extensions of random forests have been successfully employed in clinical studies [197, 198]. This research shows that correlated variables can be dealt with more effectively when elastic-net is used for regularisation. This study also showed that the regularisation technique performed better than the RSF using the AOANJRR knee data.

More studies are needed to further document the potential role of ML in clinical settings. Machine learning may be able to act as a supplementary initial screen with increased sensitivity in detecting outlier prostheses. One of the considerations is that the success of the screening process is dependent on having a pre-knowledge of clinical parameters, as this contributes significantly to the accuracy of results. In other words, the process will be compromised if some attributes that contribute to the survival of the device are not accounted for (see Appendix B). This study comprised all the recognized clinical attributes, and the most significant covariates included stability and fixation. However, other factors related to the surgeon and the subset design of devices could be investigated to study the effect of design modifications and the surgeon’s performance and experience.

Registries are recommended to use a transparent and accountable process to identify an outlier prosthesis. Aligned with the Australian standardised registry approach, time to first revision (all-cause) was the principal measure of primary joint replacement surgery. This measure is a clear indication of a problem with the primary procedure, where both the patient and surgeon have agreed that it is serious enough to require further surgical intervention [79, 199]. However, there is still scope for future studies to investigate other outcome measures for a fairer reflection of implant failure and limit the clinical end-point to specific implant-related causes. One of the principal objectives of this study was to improve the sensitivity of the initial screening rather than that of the entire identification process. Moreover, further analyses were also conducted on the prostheses identified by both the conventional and the ML approaches to determine the extent of the difference with the comparator, and the possible explanations for the variation in outcomes related to reasons for revision, type of revision, number of surgeons, etc. (detailed more in Chapter 7).

Effective feature selection techniques need to be employed for the early detection of arthroplasty devices that are at high risk of revision. Future studies can apply the proposed method to various classes of device components used for arthroplasty surgeries. The concept of prediction models to understand the significance of variables may have considerable potential to provide important context for the initial screening of prosthetic devices. The identification process developed over this comprehensive research can reduce the number of patients at risk of receiving potentially harmful devices.

6.8 Summary

The two-step feature selection may be promising as an intuitive method of outlier identification, and the Cox produced results comparable to the AOANJRR standard.

However, there was not a significant difference in the ability of the proposed techniques to detect total knee outliers.

Chapter 7. Investigations of Identified Outlier Prostheses

7.1 Overview

An exceptional and significant function of registries is that they are able to prepare population-based data on the comparative outcome of prostheses. Outcome data are required to assist an evidence-based approach to poor-performing prosthesis selection. The only source of survival outcome data for many prostheses is joint registry reports, and it is apparent that most prostheses have comparable outcomes. However, a number of them statistically have a much higher revision rate than other prostheses within the same class. The registry categorizes these prostheses as *'prostheses with higher than anticipated rates of revision'*.

The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) has the capacity to evaluate the outcome of individual prostheses or prosthesis combinations used in a surgical procedure. It seems that individual prostheses that perform well in one combination may not perform well in another combination of prosthesis construct. Therefore, the performance of a device component is relatively reliant on the prosthesis construct. The registry approach involves examining the impact of associated device components and a limited number of other confounders such as age, gender and primary diagnosis. However, it does not consider all the potential confounders simultaneously in the initial screening of total hip and knee prostheses.

The identified prostheses can be categorised according to specific groups: i) newly-identified prostheses that have been identified for the first time and are still being used; (ii) prostheses that are re-identified through the yearly statistical analysis and are still being used; and (iii) prostheses that have a much higher revision rate and are no longer available on the market. This chapter investigates the prostheses detected by all the approaches in Chapters 4 and 6. Furlong Evolution and Minimax, in addition to Score/Score and VanguardPS/Regenerex were listed in the group of re-identified devices and are still being used. However, the Emperion primary total hip replacement is no longer on the market [5].

Given the dependency of the standardised approach on the sample size and the length of follow-up time, it is becoming evident that this approach cannot as readily identify prostheses that have a delayed onset of higher rates of revision. Therefore, the registry planned to develop further strategies to solve the challenge with these poor-performing prostheses. This is why other approaches are being explored to generate comparative outcomes of primary total hip and knee prostheses.

7.2 Statistical Method

The output of our survival analyses was time-to-first-revision surgery, defined as reoperations of previous hip and knee replacements where one or more of the prosthetic components are replaced, removed, or added. The study period was 1st January 2015 to 31st December 2019. Patients with no revision or death had implant-survival times based on the time that elapsed between their initial implantation date and the end of the follow-up period. By means of the Kaplan-Meier (KM) survival analysis, five identified prostheses in primary total conventional hip and knee for OA were studied further based on their corresponding modified comparator groups developed in Chapters 3 and 5. In addition, the impact of confounding factors is examined by calculating age- and gender-adjusted hazard ratios (HRs) to check if there is a significant difference compared to the combined hazard rate of the comparator group. The unadjusted cumulative percent revision (CPR), with an accompanying 95% confidence interval (CI), was calculated after the primary surgery. This was also calculated for primary total hip and knee replacements according to each of the patient factors (i.e., age, gender, BMI, and ASA). The cumulative incidence of reasons for revisions was analysed to look at the risk associated with the most common reasons for the identified prostheses. A descriptive comparative analysis was also conducted to examine the most common types of revisions. Lastly, other potential device-related confounding impacts were investigated separately using the first stage of the AOANJRR standard. This is done by calculating revisions per 100 observed component years of the identified prostheses. The statistical analysis was performed using *R* software [158], including the packages *Survival* [159] version 3.2-11 and *Survminer* [160] version 0.4.9.

7.3 Identified Prostheses Investigation (follow-up limited to 5 years)

This section compares the identified prostheses with all other total conventional hip and knee prostheses. All the applied techniques (including Random survival and Cox regression) identified the same five prostheses as having a significantly higher rate of revision over the five-year period. These prostheses, detected by both the machine learning (ML) techniques and the standard approach, are listed in Table 7.1. Their names are given for the first time in this research as the registry reports have already reported them. Note that these devices were all detected using a greater number of observations and length of follow-up time at the time of identification.

Table 7. 1. Results for the identified prostheses detected by all the approaches.

Component	Descriptive information			AOANJRR standard		RSF	Cox
	N Revised	**N Total**	**Obs.Years**	**Stage *I***	**Stage *II***	**Minimal depth rank *Permutation P-*value**	***P*-value**
Total Hip							
Emperion	8	71	245.4	3.26 (3.01, 3.56)	4.34 (2.17, 8.68) p<0.001	2 0.009	0.009
Furlong Evolution	18	288	458.7	3.92 (3.59, 4.31)	3.28 (2.06, 5.21) p<0.001	14 0.029	<0.001
MiniMax	17	320	374.7	4.54 (4.25, 4.87)	3.02 (1.87, 4.86) p<0.001	3 0.009	0.012
Total Knee							
Score/Score	116	2648	7109.6	1.63 (1.35, 1.96)	2.21 (1.84, 2.66) p<0.001	4 P=0.009	p<0.001
VanguardPS/ Regenerex	15	222	560.1	2.68 (1.50, 4.42)	3.28 (1.97, 5.44) p<0.001	6 P=0.009	0.004

This analysis compares the total hip and knee prostheses identified as having a significantly higher revision rate with all other total prostheses correspondingly. In addition, hazard ratios are reported for the entire period to enable more specific and valid comparisons of the level of risk of revision over the entire period (Figures 7.1 to 7.5).

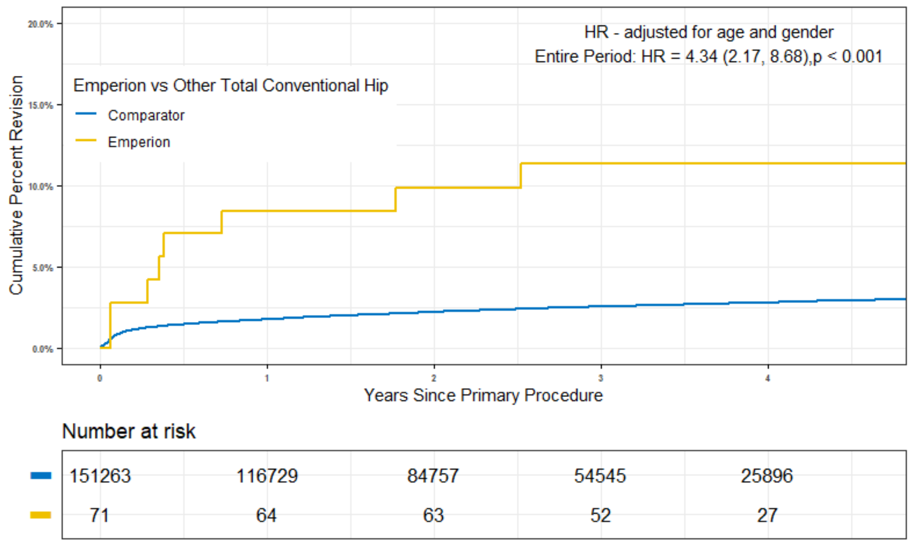
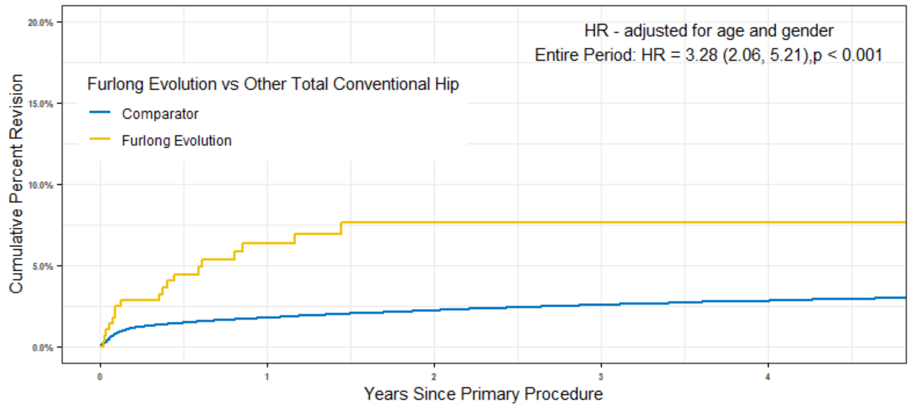


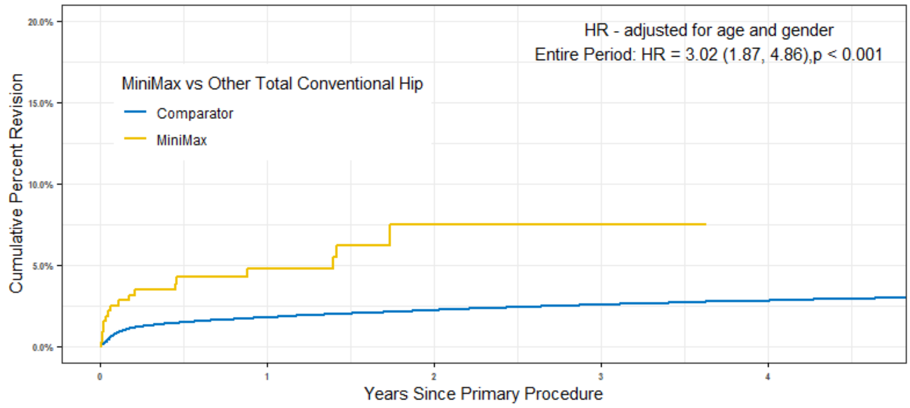
Figure 7. 1. CPR of Emperion vs modified hip comparator.



Number at risk

Comparator	151176	116788	84787	54580	25948
Furlong Evolution	288	178	88	39	28

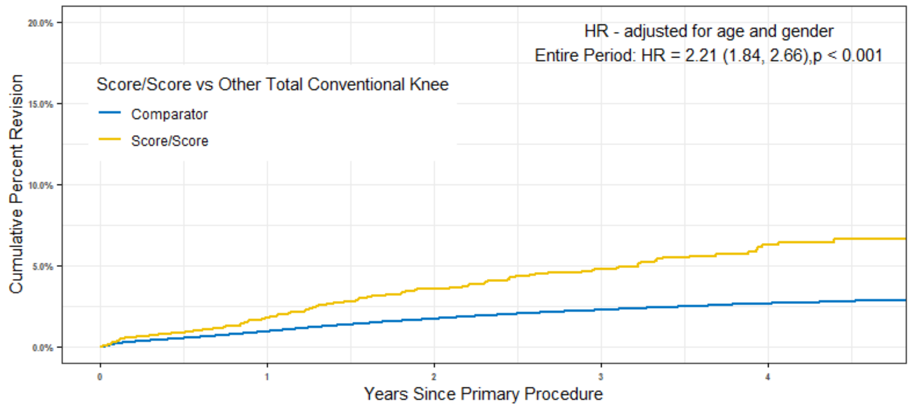
Figure 7. 2. CPR of Furlong Evolution vs modified hip comparator.



Number at risk

Comparator	151126	116839	84788	54576	25925
MiniMax	320	188	38	4	0

Figure 7. 3. CPR of MiniMax vs modified hip comparator.



Number at risk

Comparator	241180	188640	137546	88473	43000
Score/Score	2648	2180	1730	1198	639

Figure 7. 4. CPR of Score/Score vs conventional knee comparator.

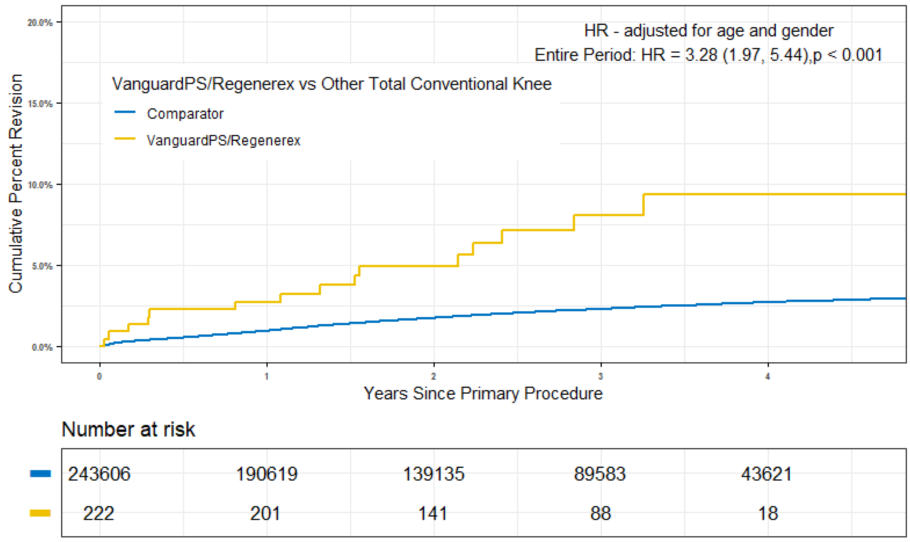


Figure 7. 5. CPR of Vanguard PS/Regenerex vs conventional knee comparator.

7.4 Reason for Revision

The competing risk of reasons for revisions is shown in Figures 7.6 to 7.9. These results can be used to estimate the risk of requiring revision for each of the listed diagnoses such as infection, loosening, prosthesis dislocation, and fracture. Different outcomes for the reasons for revision with the same distribution of follow-up time may identify problems of concern. Given the criterion that a diagnosis should require more than two revisions, the most common revision diagnosis is dislocation for the Emperion femoral stem, infection for Furlong Evolution, and fracture for MiniMax primary total hip outliers. Infection appears to be the most probable cause of revision for both of the identified primary total knee outliers. The figures below detail the cumulative incidence of the most common reasons for revisions. The most common reasons for revision are included if each of these reasons accounts for more than two procedures. For example, two of the outlier prostheses have only one revision diagnosis according to the criterion.

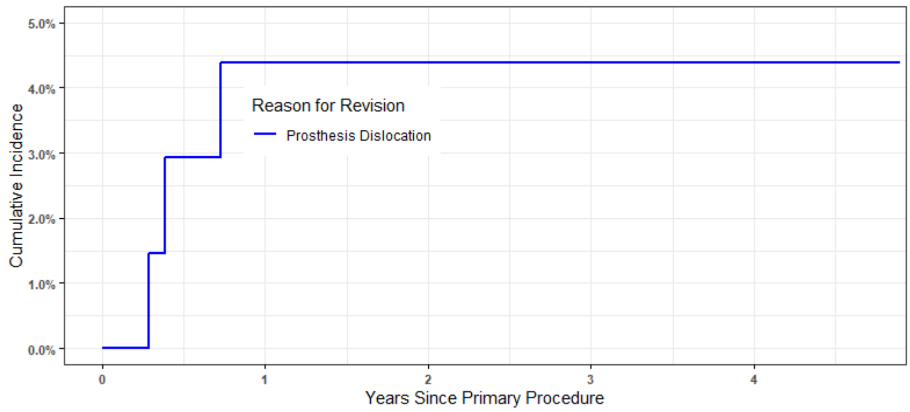


Figure 7. 6. Cumulative incidence of most common revision diagnosis for Emperion.

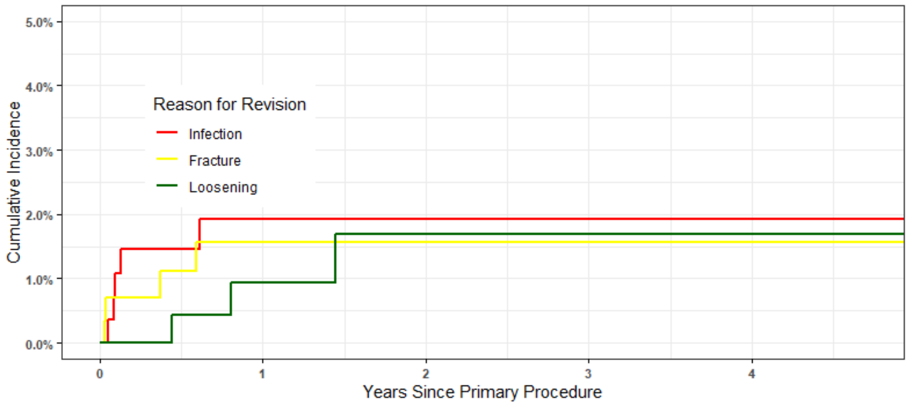


Figure 7. 7. Cumulative incidence of most common revision diagnosis for Furlong Evolution.

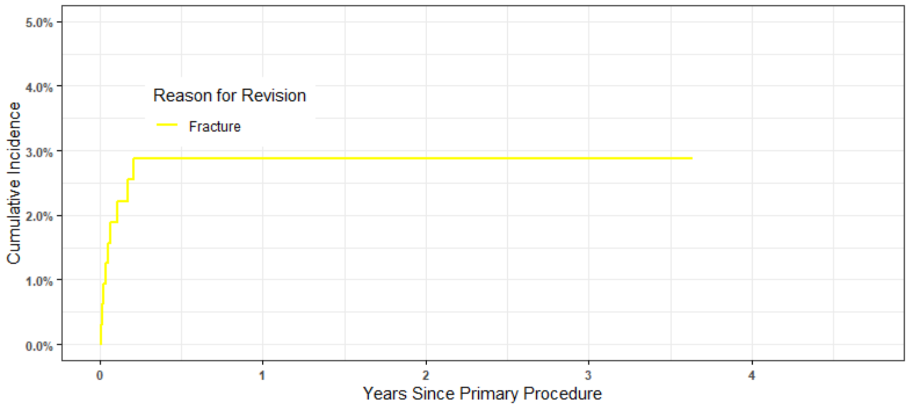


Figure 7. 8. Cumulative incidence of most common revision diagnosis for MiniMax.

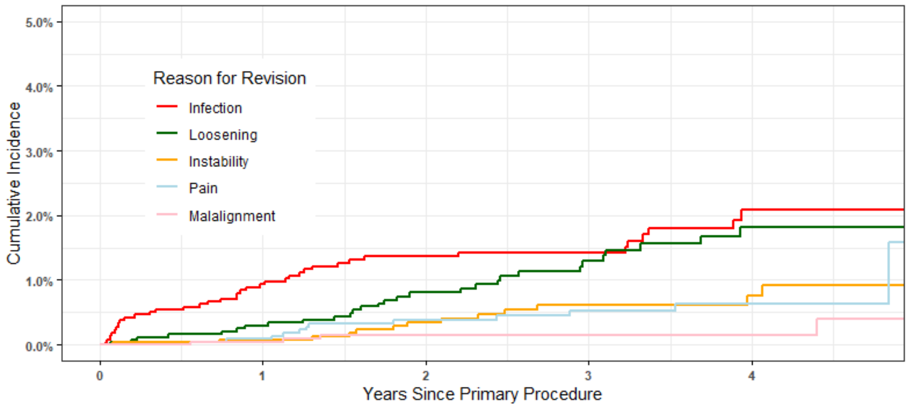


Figure 7. 9. Cumulative incidence of most common revision diagnosis for Score/Score.

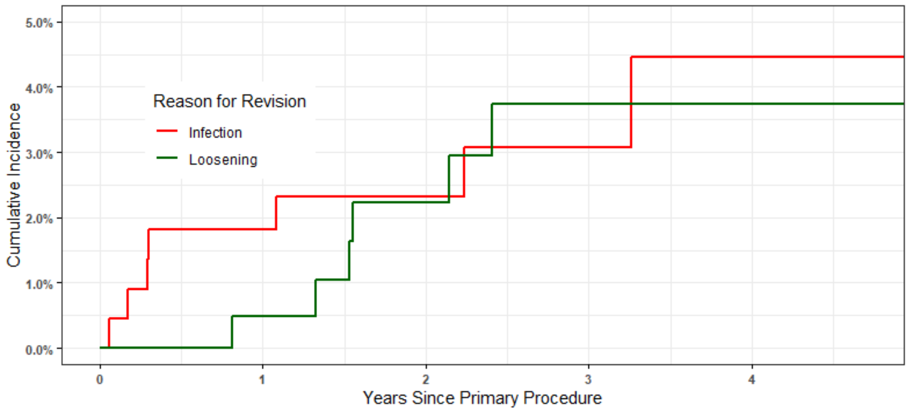


Figure 7. 10. Cumulative incidence of most common revision diagnosis for Vanguard PS/Regenerex.

7.5 Type of Revision

This section discusses the type of revision or the components responsible for revising the total hip and knee outlier prostheses. This part undertook this comparison to determine whether one or more of the components that are replaced differ from the components replaced in revisions of the corresponding comparator groups. The most common major types of revisions were the femoral component for the hip outliers and 'TKR (tibial/femoral)' for the knee outlier prostheses. In addition, the hip prostheses had minor revisions for 'head/insert' or 'head only', and the knee outliers for 'insert only'. These descriptive analyses show the same outcomes for the corresponding comparator groups.

Table 7. 2. Type of revision performed for primary total conventional hip replacement.

	Emperion		Modified hip comparator	
Type of Revision	**Number**	**Percent**	**Number**	**Percent**
Femoral Component	2	25.0	1076	31.7
Acetabular Component	-	-	569	16.8
THR (Femoral/Acetabular)	1	12.5	290	8.5
Cement Spacer	-	-	102	3.0
Removal of Prostheses	-	-	22	0.6
Reinsertion of Components	-	-	9	0.3
Bipolar Head and Femoral	-	-	1	0.0
N Major	**3**	**37.5**	**2069**	**61.2**
Head/Insert	4	50.0	995	29.3
Head Only	1	12.5	216	6.4
Minor Components	-	-	57	1.7
Insert Only	-	-	45	1.3
N Minor	**5**	**62.5**	**1313**	**38.8**
Total	**8**	**100.0**	**3382**	**100.0**

Table 7. 3. Type of revision performed for primary total conventional hip replacement.

	Furlong Evolution		Modified hip comparator	
Type of Revision	**Number**	**Percent**	**Number**	**Percent**
Femoral Component	7	38.9	1071	31.8
Acetabular Component	4	22.2	565	16.7
THR (Femoral/Acetabular)	2	11.1	289	8.6
Cement Spacer	2	11.1	100	3.0
Removal of Prostheses	-	-	22	0.6
Reinsertion of Components	-	-	9	0.3
Bipolar Head and Femoral	-	-	1	0.0
N Major	**15**	**83.3**	**2057**	**61.0**
Head/Insert	2	11.1	997	29.6
Head Only	1	5.5	216	6.4
Minor Components	-	-	57	1.7
Insert Only	-	-	45	1.3
N Minor	**3**	**16.7**	**1315**	**39.0**
Total	**18**	**100.0**	**3372**	**100.0**

Table 7. 4. Type of revision performed for primary total conventional hip replacement.

	MiniMax		Modified hip comparator	
Type of Revision	**Number**	**Percent**	**Number**	**Percent**
Femoral Component	9	52.94	1069	31.7
Acetabular Component	1	5.88	568	16.8
THR (Femoral/Acetabular)	-	-	291	8.6
Cement Spacer	1	5.88	101	3.0
Removal of Prostheses	-	-	22	0.6
Reinsertion of Components	1	5.88	8	0.2
Bipolar Head and Femoral	-	-	1	0.0
N Major	**12**	**70.6**	**2060**	**61.1**
Head/Insert	2	11.76	997	29.5
Head Only	3	17.65	214	6.3
Minor Components	-	-	57	1.7
Insert Only	-	-	45	1.3
N Minor	**5**	**29.4**	**1313**	**38.9**
Total	**17**	**100.0**	**3373**	**100.0**

Table 7. 5. Type of revision performed for primary total conventional knee replacement.

	Score/Score		Modified conventional knee comparator	
Type of Revision	**Number**	**Percent**	**Number**	**Percent**
TKR (Tibial/Femoral)	50	43.1	890	20.5
Tibial Component	4	3.4	305	7.0
Cement Spacer	8	6.9	234	5.4
Femoral Component	4	3.4	223	5.1
Removal of Prostheses	-	-	31	0.7
Total Femoral	-	-	3	0.1
Reinsertion of Components	-	-	1	0.0
N Major	**66**	**56.9**	**1687**	**38.8**
Insert Only	33	28.4	1710	39.3
Patella Only	15	12.9	637	14.6
Insert/Patella	2	1.7	302	6.9
Minor Components	-	-	9	0.2
Cement Only	-	-	3	0.1
N Minor	**50**	**43.1**	**2661**	**61.2**
Total	**116**	**100.0**	**4,348**	**100.0**

Table 7. 6. Type of revision performed for primary total conventional knee replacement.

	Vanguard PS/Regenerex		Modified conventional knee comparator	
Type of Revision	**Number**	**Percent**	**Number**	**Percent**
TKR (Tibial/Femoral)	7	46.7	933	21.0
Tibial Component	-	-	309	6.9
Cement Spacer	3	20.0	239	5.4
Femoral Component	2	13.3	225	5.0
Removal of Prostheses	-	-	31	0.7
Total Femoral	-	-	3	0.1
Reinsertion of Components	-	-	1	0.0
N Major	**12**	**80.0**	**1741**	**39.1**
Insert Only	3	20.0	1740	39.1
Patella Only	-	-	652	14.6
Insert/Patella	-	-	304	6.8
Minor Components	-	-	9	0.2
Cement Only	-	-	3	0.1
N Minor	**3**	**20.0**	**2708**	**60.9**
Total	**15**	**100.0**	**4,449**	**100.0**

7.6 Prosthesis-related Confounding Factors

Revision Rates of Outlier Prostheses by Fixation

Tables 7.7 to 7.11 present revision rates of primary total hip and knee outliers by fixation as several prostheses have more than one option for fixation. Moreover, prostheses with an alternative fixation may be used by surgeons regardless of the recommended approach (e.g., a cementless prosthesis that is cemented or cementless). Total hip outliers were used only with a recommended cementless fixation. However, there are variations in the outcome of total knee outliers by fixation. The hybrid fixation was the most chosen option by surgeons for the total knee outliers.

Table 7. 7. Revision rates of Emperion primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	8	71	245.4	3.26 (3.01, 3.56)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 7. 8. Revision rates of Furlong Evolution primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	18	288	458.7	3.92 (3.59, 4.31)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Table 7. 9. Revision rates of MiniMax primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	17	320	374.7	4.54 (4.25, 4.87)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Table 7. 10. Revision rates of Score/Score primary total knee replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cemented	14	497	1366.9	1.02 (0.56, 1.72)
Cementless	50	932	2390.8	2.09 (1.55, 2.76)
Hybrid (Tibial cemented)	51	1213	3330.4	1.53 (1.14, 2.01)
Hybrid (Tibial cementless)	1	6	21.5	4.65 (0.12, 25.91)
Total	**116**	**2648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 7. 11. Revision rates of Vanguard PS/Regenerex primary total knee replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cemented	0	2	4.8	0.00 (0.00, 76.53)
Cementless	2	48	154	1.30 (0.16, 4.69)
Hybrid (Tibial cementless)	13	172	401.2	3.24 (1.72, 5.54)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Revision Rates of Outlier Prostheses by Bearing Surface

This analysis was conducted because some prostheses had a combination of various bearing surfaces. All bearing surfaces used with these outlier prostheses are listed below in Tables 7.12 to 7.16. The investigation of the revision rates according to the bearing surface indicates the role of material designs in the performance of prostheses. This information is expected to lead to better long-term outcomes for prostheses with modern bearing surfaces (including ceramic/ceramic and the femoral head materials in conjunction with XLPE) [112, 200]. However, this outcome was not obtained through our comparative analyses because of a significant sample size/follow-up variation. The limited five-year follow-up may be a constraint preventing the accurate reflection of the difference between the bearing couples.

Table 7. 12. Revision rates of Emperion primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	4	54	191.2	2.09 (0.57, 5.35)
Ceramicised Metal/XLPE	3	12	37.6	7.97 (1.64, 23.30)
Metal/XLPE	1	5	16.5	6.06 (0.15, 33.79)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 7. 13. Revision rates of Furlong Evolution primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	10	142	235.8	4.24 (2.03, 7.80)
Ceramic/Non-XLPE	7	114	129.6	5.40 (2.17, 11.13)
Ceramic/XLPE	0	13	23.9	0.00 (0.00, 15.46)
Ceramic/XLPE + Antioxidant	0	4	12.3	0.00 (0.00, 29.97)
Metal/Non-XLPE	1	15	57.2	1.75 (0.04, 9.74)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Table 7. 14. Revision rates of MiniMax primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	8	163	161.6	4.89 (2.14, 9.75)
Ceramic/Non-XLPE	5	111	179.2	2.79 (0.90, 6.51)
Ceramic/XLPE	3	43	31.8	9.43 (1.94, 27.57)
Metal/Non-XLPE	0	1	0.2	0.00 (0.00, 1676.76)
Metal/XLPE	1	2	1.8	55.55 (1.41, 309.53)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Table 7. 15. Revision rates of Score/Score primary total conventional knee replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Non-XLPE	116	2648	7109.6	1.63 (1.35, 1.96)
Total	**116**	**2648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 7. 16. Revision rates of Vanguard PS/Regenerex primary total conventional knee replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Non-XLPE	15	218	543.1	2.76 (1.54, 4.55)
XLPE + Antioxidant	0	4	17	0.00 (0.00, 21.75)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Revision Rates of Outlier Prostheses by Surgeon *ID*

The surgeon *ID*s were used to determine the number of surgeons who used the outlier prostheses in the primary total joint procedures. The aim was to find whether there was a correlation between surgeons and the outlier prostheses they used. To the best of our information, this is the first research designed to investigate the role of a single surgeon in outlier identification. Results show that Furlong Evolution, MiniMax, Score/Score, and Vanguard PS/Regenerex have a higher-than-expected rate of revisions for most surgeons (Tables 7.17-7.21). This can be an indication of the poor-performing prostheses regardless of the expertise or experience of the surgeon. However, the interaction between the Emperion femoral stem and the surgeons needs further analysis as this hip prosthesis had been used by only one surgeon (685).

Table 7. 17. Revision rates of Emperion primary total conventional hip replacement by surgeon id.

Surgeon ID	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
685	4	53	188.3	2.12 (0.58, 5.44)

Table 7. 18. Revision rates of Furlong Evolution primary total conventional hip replacement by surgeon id.

Surgeon ID	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
587	1	19	38.4	2.6 (0.6, 14.5)
1246	2	17	26.8	7.46 (0.9, 26.95)
1357	1	27	105.4	0.95 (0.2, 5.29)
1421	0	19	9	0.00 (0.00, 41.13)
1726	1	61	56.9	1.76 (0.04, 9.79)
1745	9	112	144.4	6.23 (3.32, 12.74)

Table 7. 19. Revision rates of Minimax primary total conventional hip replacement by surgeon id.

Surgeon ID	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
804	2	93	94.2	2.12 (0.26, 7.67)
1041	5	110	173.7	2.88 (0.93, 6.72)
1195	2	17	5.1	39.19 (4.75, 141.57)
1421	1	22	29.1	3.43 (0.09, 19.14)
1529	2	20	28.7	6.96 (0.84, 25.14)
1717	0	10	13	0.00 (0.00, 28.36)
1914	1	13	6	16.68 (0.42, 92.92)

Table 7. 20. Revision rates of Score/Score primary total knee replacement by surgeon id.

Surgeon ID	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
57	1	28	125.5	0.8 (0.02, 4.44)
153	1	23	89.4	1.12 (0.03, 6.23)
173	1	45	146.3	0.68 (0.02, 3.81)
275	4	143	516.5	0.77 (0.21, 1.98)
282	6	152	618	0.97 (0.36, 2.11)
456	0	10	12.3	0.00 (0.00, 30.09)
495	9	191	623.9	1.44 (0.66, 2.74)
895	3	85	332.9	0.9 (0.18, 2.63)
934	0	11	34.3	0.00 (0.00, 10.75)
961	5	159	243.6	2.05 (0.67, 4.79)
1009	2	58	239.5	0.83 (0.10, 3.02)
1119	34	511	1637.8	2.07 (1.44, 2.9)
1149	16	379	716.9	2.23 (1.27, 3.62)
1177	15	197	649.9	2.31 (1.29, 3.81)
1195	0	38	111.6	0.00 (0.00, 3.30)
1218	7	439	414.4	1.69 (0.68, 3.48)
1258	0	17	64.6	0.00 (0.00, 5.71)
1745	12	106	374.5	3.2 (1.65, 5.6)
1810	0	20	72.1	0.00 (0.00, 5.11)

Table 7. 21. Revision rates of Vanguard PS/Regenerex primary total knee replacement by surgeon id.

Surgeon ID	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
253	2	42	135.9	1.47 (0.18, 5.32)
660	13	172	401.2	3.24 (1.72, 5.54)

The criterion that was taken into account was twice the modified comparator over a five-year follow-up. This can also be modified to 1.5 times the comparator to better reflect the role of surgeons when using a specific hip or knee outlier. Results show that the same surgeons (1195 and 1745) used most of the outlier prostheses and provided significant interactions on Furlong Evolution, MiniMax and Score/Score performance. These surgeons recorded revision rates much higher than twice the comparator and even higher than most of the other surgeons who used the same device components, such as for the MiniMax primary total hip replacement.

Number of Total and Revision of Outlier Prostheses by Year of Implant

Table 7.22 shows the number of primary joint procedures performed annually using the outlier prostheses and the number revised. There is less follow-time for the primary operations performed in later years; consequently, the number revised is estimated to be less than the number of revisions in earlier years. For example, a primary procedure performed in 2018 has a maximum of two years to be revised although a primary procedure performed in 2016 has a maximum of four years until a potential revision surgery.

Table 7. 22. Number of total and revision of total hip outliers by year of implant.

	Emperion		Furlong Evolution		MiniMax	
Year of implant	N Revised	N Total	N Revised	N Total	N Revised	N Total
2015	4	32	2	31	0	0
2016	4	29	0	11	0	4
2017	0	10	3	52	3	37
2018	0	0	7	94	6	155
2019	0	0	6	100	8	124
Total	**8**	**71**	**18**	**288**	**17**	**320**

	Score/Score		Vanguard PS/Regenerex	
Year of implant	N Revised	N Total	N Revised	N Total
2015	36	703	0	18
2016	34	579	5	76
2017	22	527	10	58
2018	19	419	0	56
2019	5	420	0	14
Total	**116**	**2648**	**15**	**222**

Revision Rates of Total Hip Outlier Prostheses by Associated Component

An individual total hip prosthesis may be combined and used with several components. This analysis has been conducted to investigate whether the revision rate varies according to the combined component. This part of the analysis was conducted to investigate the individual hip outliers (femoral stems) combined with acetabular cups. Tables 7.23-7.25 show that the issue could be related to outlier femoral stems rather than the matched acetabular cups. The higher-than-expected revision rate is not significantly correlated with the performance of acetabular cups. The Emperion femoral stem was mainly combined with 'R3' cup, Furlong Evolution with the Furlong cup, and Minimax with a broader range of cups such as 'Mpact', 'Versafitcup CC', and 'Versafitcup DM'.

Table 7. 23. Revision rates of Emperion primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
R3	8	70	240.6	3.32 (1.43, 6.55)
Trabecular Metal (Shell)	0	1	4.8	0.00 (0.00, 77.17)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 7. 24. Revision rates of Furlong Evolution primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Acetabular Shell (Global)	0	4	17.5	0.00 (0.00, 21.11)
Adaptive	0	4	12.3	0.00 (0.00, 29.97)
Delta-TT	0	2	4.3	0.00 (0.00, 85.99)
Furlong	18	241	381.7	4.71 (2.79, 7.45)
Logical G	0	14	24.6	0.00 (0.00, 14.97)
Mpact	0	18	9.1	0.00 (0.00, 40.63)
PINNACLE	0	1	0.2	0.00 (0.00, 1676.76)
Trident/Tritanium (Shell)	0	1	3.4	0.00 (0.00, 109.46)
Versafitcup CC	0	3	5.7	0.00 (0.00, 65.06)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Table 7. 25. Revision rates of MiniMax primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Agilis	0	1	0.2	0.00 (0.00, 1756.61)
Mpact	8	123	99.5	8.04 (3.47, 15.85)
Trinity	0	1	3.4	0.00 (0.00, 106.92)
Versafitcup CC	4	87	98.9	4.05 (1.10, 10.36)
Versafitcup DM	5	108	172.7	2.89 (0.94, 6.76)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Revision Rates of Total Hip Outlier Prostheses by Head Size

To assess the effect of head size on the prosthesis survillance, an analysis was conducted to compare head size groups (<32mm versus ≥32mm) for the total hip outliers. The Emperion femoral stem with head sizes ≥32mm has a higher revision rate than the other constructs of this device with smaller head sizes. Larger head sizes showed better outcomes for Furlong Evolution, although our sample size was less than 10 for both Emperion and Furlong Evolution primary total hip replacements (Tables 7.26 and 7.27). However, in the case of Minimax, there was a higher revision rate for larger head sizes over the 5-years follow-up (Table 7.28). The results illustrate that the outcomes of outlier prostheses vary according to head size, although better outcomes are expected for larger head sizes as there could be a lower risk of early dislocation [153, 157].

Table 7. 26. Revision rates of Emperion primary total conventional hip replacement by head size.

Patella used	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
< 32	0	4	14.9	0.00 (0.00, 24.77)
>= 32	8	67	230.5	3.47 (1.50, 6.84)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 7. 27. Revision rates of Furlong Evolution primary total conventional hip replacement by head size.

Patella used	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
< 32	1	6	11.4	8.77 (0.22, 48.87)
>= 32	17	282	447.3	3.80 (2.21, 6.08)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Table 7. 28. Revision rates of MiniMax primary total conventional hip replacement by head size.

Patella used	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
< 32	5	113	179.7	2.78 (0.90, 6.49)
>= 32	12	207	194.9	6.16 (3.18, 10.75)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Revision Rates of Total Knee Outlier Prostheses by Patella Usage

According to the AOANJRR reports, primary total knee procedures with patellar component have less overall revision rate than procedures without patellar resurfacing [201, 202]. This is also indicated by the results provided in Table 7.29 for Score/Score primary total conventional knee replacement. However, when the patella is resurfaced, the revision rate of Vanguard PS/Regenerex was much higher than that of with no patella (Table 7.30). Overall, patellar resurfacing effects vary dependent on the outlier prosthesis used, as other factors can potentially contribute to the outcome.

Table 7. 29. Revision rates of Score/Score primary total knee replacement by patella usage.

Patella used	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
YES	97	2232	6210.5	1.56 (1.27, 1.90)
NO	19	416	899	2.11 (1.27, 3.30)
Total	**116**	**2648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 7. 30. Revision rates of Vanguard PS/Regenerex primary total knee replacement by patella usage.

Patella used	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
YES	15	200	495.2	3.03 (1.69, 5.00)
NO	0	22	64.9	0.00 (0.00, 5.69)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Revision Rates of Total Knee Outlier Prostheses by Bearing Mobility

Revision rates of primary total knee outliers by bearing mobility were calculated as some prosthesis constructs are combined with various bearing mobilities. All the mobilities used with these two outlier combinations are listed in Tables 7.31-7.32. Fixed-bearing prostheses include non-modular tibial knee constructs and those prostheses with fixed inserts without the ability to move relative to the baseplate. According to the registry report, fixed designs have a lower revision rate than do mobile designs during the first seven years [5]. However, the identified outliers were only used by one specific mobility design.

Table 7. 31. Revision rates of Score/Score primary total knee replacement by bearing mobility.

Bearing mobility	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Rotating	116	2648	7109.6	1.63 (1.35, 1.96)
Total	**116**	**2648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 7. 32. Revision rates of Vanguard PS/Regenerex primary total knee replacement by bearing mobility.

Bearing mobility	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Fixed	15	222	560.1	2.68 (1.50, 4.42)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Revision Rates of Total Knee Outlier Prostheses by Stability Design

This analysis was conducted because some prostheses are combined with various stability designs. All stabilities used with these outlier combinations are listed in Tables 7.33 and 7.34. Stability denotes specific prosthetic features proposed to substitute for the intrinsic stability of knee ligaments. The five groups are: minimally stabilised, medial pivot design, posterior stabilised, fully stabilised, and hinged prostheses. The two knee outliers were manufactured only by standard stability; Minimally stabilised was the design used for 'Score/Score' and posterior stabilised for the 'Vanguard PS/Regenerex'.

Table 7. 33. Revision rates of Score/Score primary total knee replacement by stability.

Stability	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Minimally Stabilised	116	2648	7109.6	1.63 (1.35, 1.96)
Total	**116**	**2648**	**7109.6**	**1.63 (1.35, 1.96)**

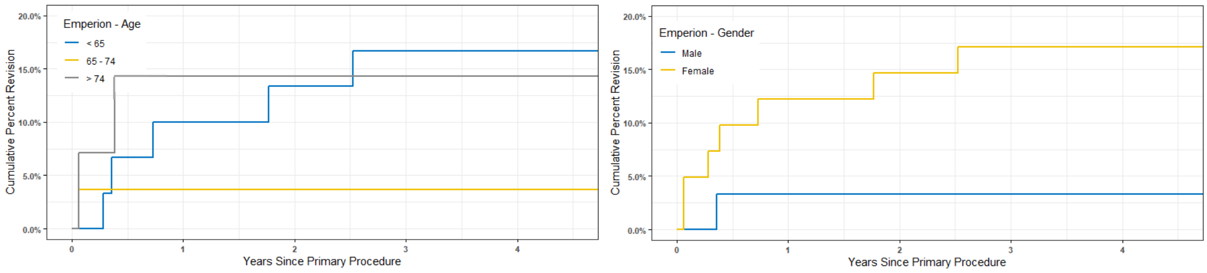
Table 7. 34. Revision rates of Vanguard PS/Regenerex primary total knee replacement by bearing mobility.

Bearing mobility	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Posterior Stabilised	15	222	560.1	2.68 (1.50, 4.42)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

7.7 Patient Characteristics

CPR of Outlier Prostheses by Age and Gender

The identified femoral stems were most commonly implanted for primary total hip replacement in patients younger than 65-years (Figures 7.11-7.13). This is also more common for patients aged 65 to 74, where outlier prosthesis combinations were mainly used in primary total conventional knee surgeries (Figures 7.14 and 7.15). In the short term (less than three months), older patients who underwent an outlier total hip replacement had a higher rate of revision, while this risk is higher in younger patients over the long term. This is because younger patients could potentially engage in a higher level of routine activities that increase revision risk. When the Furlong Evolution stem or the Score/Score were used, female patients showed a higher number at risk with lower overall CPR compared to male patients. However, female patients showed higher rates of revision than males for the other outlier prostheses, including Emperion, MiniMax, and Vanguard PS/Regenerex.



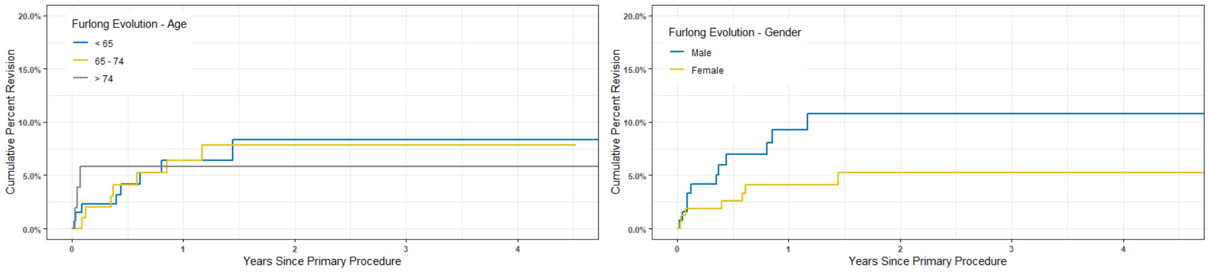
Number at risk

Age	0	1	2	3	4
< 65	30	27	26	20	10
65 - 74	27	26	26	21	14
> 74	14	11	11	11	3

Number at risk

Gender	0	1	2	3	4
Male	30	28	28	25	15
Female	41	36	35	27	12

Figure 7. 11. CPR of Emperion primary total conventional hip by age and gender.



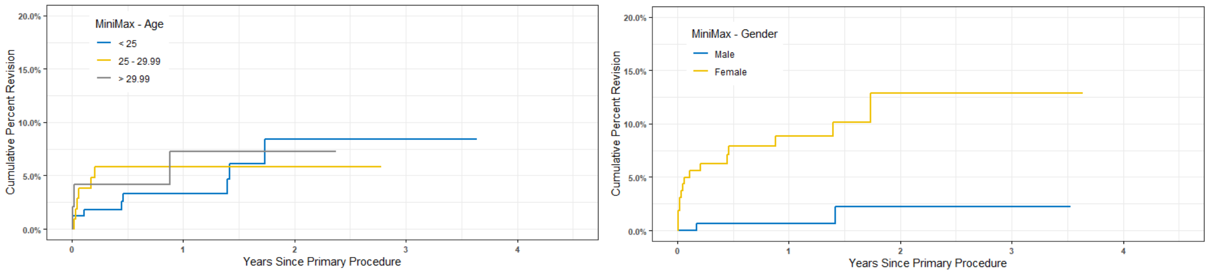
Number at risk

Age	0	1	2	3	4
< 65	135	66	26	17	13
65 - 74	101	75	41	12	8
> 74	52	37	21	10	7

Number at risk

Gender	0	1	2	3	4
Male	125	68	33	14	12
Female	163	110	55	25	16

Figure 7. 12. CPR of Furlong Evolution primary total conventional hip by age and gender.



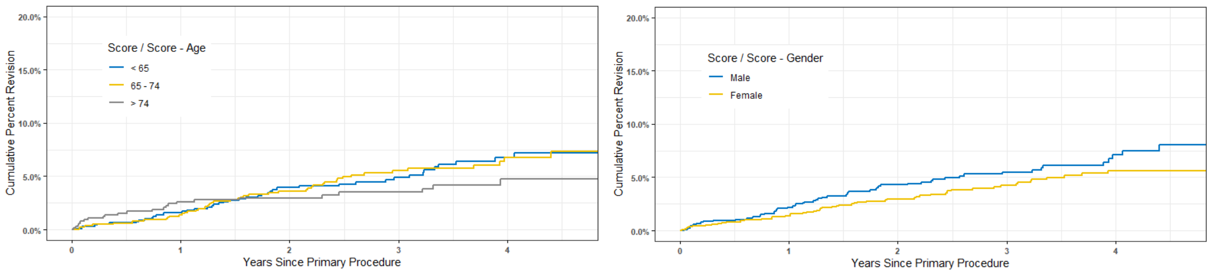
Number at risk

Age	0	1	2	3	4
< 25	167	100	21	4	0
25 - 29.99	105	60	12	0	0
> 29.99	48	28	5	0	0

Number at risk

Gender	0	1	2	3	4
Male	159	97	20	3	0
Female	161	91	18	1	0

Figure 7. 13. CPR of MiniMax primary total conventional hip by age and gender.



Number at risk

Age	0	1	2	3	4
< 65	955	783	619	427	230
65 - 74	1035	855	688	466	249
> 74	658	542	423	305	160

Number at risk

Gender	0	1	2	3	4
Male	1202	967	749	511	267
Female	1446	1213	981	687	372

Figure 7. 14. CPR of Score/Score primary total conventional knee by age and gender.

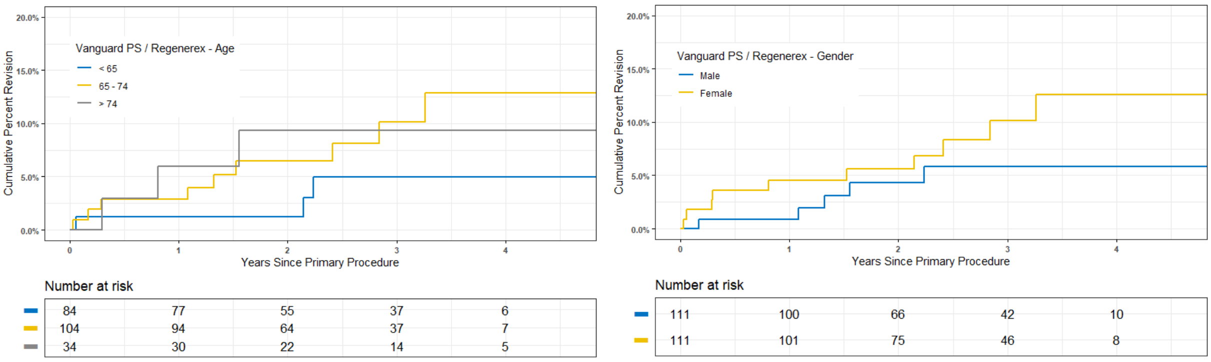


Figure 7. 15. CPR of Vanguard PS/Regenerex primary total conventional knee by age and gender.

CPR of Outlier Prostheses by BMI and ASA Score

The CPR rates of the identified prostheses according to the classified levels of Body Mass Index (BMI) and the American Society of Anesthesiologists (ASA) are shown in Figures 7.16-7.20. Patients with ASA scores less than 3, and BMI equal to or greater than 30, usually had operations using outlier prostheses identified. Except for the Emperion femoral stem, no difference is observed in the CPRs for outlier prostheses by ASA scores (Figures 7.16-7.20). For the hip and knee replacements where poorly-performing prostheses were used, most procedures involved patients with high levels of obesity. Overall, there was a difference in the outcome of outlier prostheses when the classified BMI values were compared. Although our understanding of patient-reported knee and hip replacement outcomes has progressed, it still needs to be refined. National registries are helping us learn more about knee replacements, and new statistical approaches should be used to derive the most from collected data.

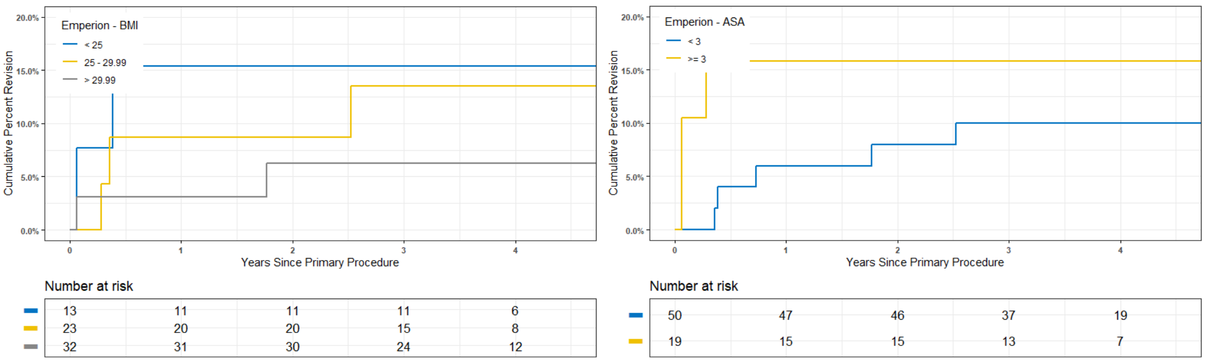


Figure 7. 16. CPR of Emperion primary total conventional hip by BMI and ASA score.

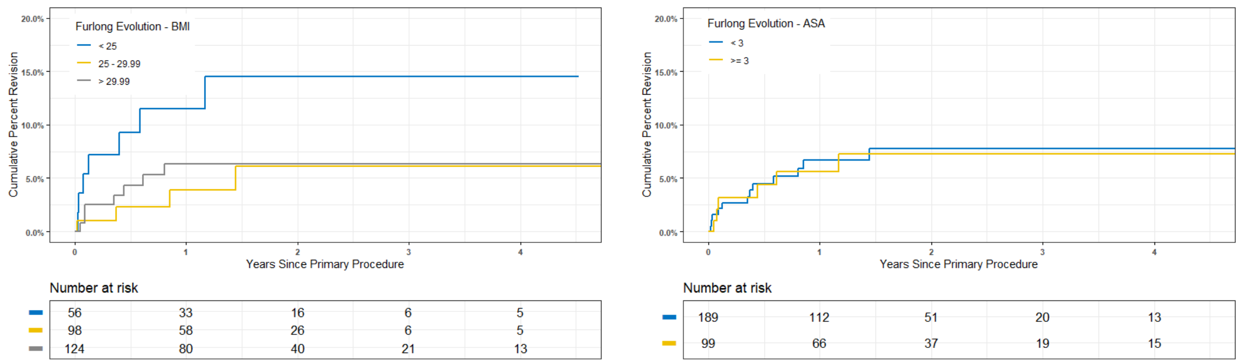


Figure 7. 17. CPR of Furlong Evolution primary total conventional hip by BMI and ASA score.

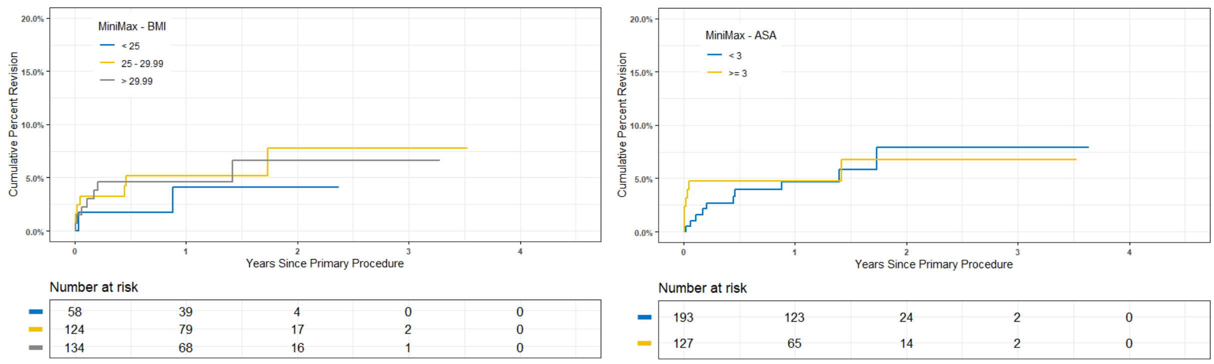


Figure 7. 18. CPR of MiniMax primary total conventional hip by BMI and ASA score.

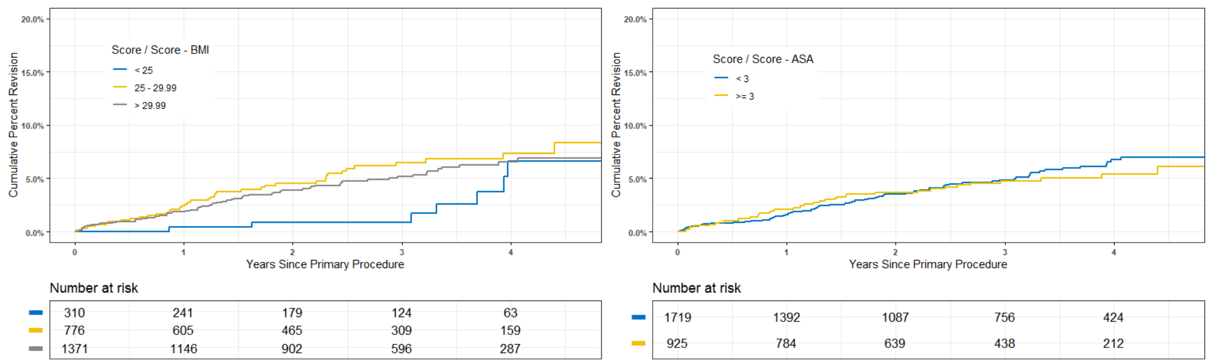


Figure 7. 19. CPR of Score/Score primary total conventional knee by BMI and ASA score.

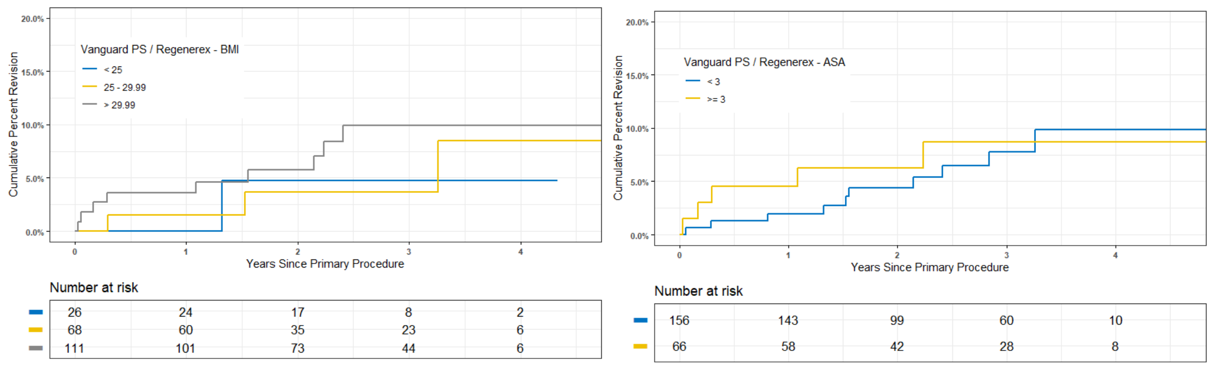


Figure 7. 20. CPR of Vanguard PS/Regenerex primary total conventional knee by BMI and ASA score.

7.8 Discussion

This study analysed the three poor-performing total hip femoral stems and two identified total knee prosthesis combinations. Emperion, Furlong Evolution, and MiniMax were detected by both the ML techniques as well as the standard approach. This was done by utilising the modified comparator groups developed in Chapters 3 and 5 and by taking into account the confounding presented in Chapters 4 and 6. Non-contemporary practices have been excluded from the comparator for all primary total conventional hip studies. Complex procedures were excluded from the comparator group of total conventional knee prostheses. The revision rates of the detected prostheses exceeded twice that of the modified comparator groups. In addition, the proposed ML techniques detected these outliers, taking into account the four available patient factors and prosthesis-related confounding factors.

Many factors may impact the effectiveness of joint replacement surgery over time. Joint replacement registries are able to detect differences in outcomes based on patient-, surgery-, or prosthesis-specific factors [203-206]. The comparative analyses presented in this chapter were designed to study the outlier prostheses, taking into account the potential confounding. Arthroplasty registries critically play a role in preparing post-market surveillance, helping practitioners to comprehend prosthetic use and positively affect patient outcomes [81, 207, 208]. Registries should use a precise and reliable approach to identify an outlier. This chapter was an attempt to evaluate the extent of confounding and determine the possible impacts of confounding on outlier prosthesis survival. The details regarding the prostheses with higher-than-expected revision rates are far less widely reported. Therefore, there is a need to investigate the various device- and non-device-related factors that may clinically affect the survival outcome.

This research showed that routine screening might require a specific post-analysis that includes more confounding factors. In addition, ML is able to assist the process by considering more salient factors and evaluating their interactions. The registry will improve the identification approach progressively, considering that decreasing the number of procedures available may adversely affect statistical accuracy. Two different statistical methods were used for this comparative post-analysis with respect to the available number of observations. The component year formula is able to provide better comparative outcomes of study populations when there are limitations related to sample sizes. However, a more precise rate of revision was obtained by undertaking KM survival analyses for a number of confounding factors with reasonable sample sizes and follow-ups.

7.9 Summary

In this chapter, the performance of outlier prostheses was studied in terms of clinically-known confounding factors to determine the impact of design- and patient-related variables on the performance of outlier prostheses. The results showed variations in the outcome of identified total hip and knee outliers in regard to bearing surfaces. In addition, the longevity of total knee outliers was associated with the method of fixation and the patella that was used. There was also a significant variation in the surveillance of total hip prostheses with the femoral head size and the associated device component.

Chapter 8. Discussion

8.1 Joint Registry Approach for the Identification of Outlier Prostheses

Arthroplasty registries differ in their approaches to identifying outlier prostheses within the community. The Swedish hip arthroplasty register circulates survivorship curves but there is no specific comparison of the prostheses performance [170]. The Norwegian registry documents the surveillance of prostheses and distributes outcomes in peer-reviewed journals, but the annual report does not include any particular survivorship curves [209]. The New Zealand registry publishes tables of prostheses outcomes but does not identify outlier prostheses [40]. The National Joint Registry for England and Wales has developed a subcommittee to discuss unique strategies for each prosthesis [5]. The Swedish Knee Arthroplasty Register uses an implant as a reference when comparing the outcomes of other prostheses [34]. The main challenge of all these approaches is to identify the most appropriate comparator for a liable comparative study of prostheses performance.

A number of devices are identified annually as poor-performing prostheses given the post-surgery outcomes. Joint registries (JRs) need to apply appropriate and effective methods to recognize poor-performing prostheses and evaluate all the potential variables impacting the monitoring process of total joint prostheses. JRs can identify variations in the outcome by estimating time to first revision using Kaplan-Meier survival analyses [203-206]. The variation in the performance of prostheses indicates the necessity for thorough pre-market evaluation and careful post-market assessment.

Joint registries assist the community in understanding the use of prosthetic devices and improve patient outcomes by providing high-quality post-market survival results [207, 208]. Individuals or prosthesis combinations are regularly identified if they indicate a much higher revision rate than others within the same broad class [210, 211]. Outlier detection will continue to improve with the application of more effective strategies. Many approaches have been investigated for systematically reducing the revision rate but the details were far less widely reported. International collaborations can improve the identification by sharing data and assessing the surveillance of prostheses using a more extensive database.

The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) determines the possible reasons for higher-than-expected revision rates of outlier prostheses. The registry developed a standardized multi-stage approach with an initial screening analysis that automatically identifies an outlier individually or in combination with the associated components. This occurs when the revision rate per 100 component years exceeds twice that of all others [212]. Then, a limited number of potential confounding factors (including age, gender, and primary diagnosis) are reviewed as a part of this early assessment [15].

The registry assesses patient-, surgeon- and device-related variables that may significantly contribute to the observed higher revision rate. For example, some prostheses may be detected by the initial screening test but not be reported for various reasons such as inadequate sample size, their use in complex primary situations or when combined with a poor-performing component. However, there are several limitations with the standardised approach as it is not able to consider all the known confounding factors simultaneously. The method has significant limitations when handling interactions of highly-correlated variables.

Joint replacement registries are able to identify differences in such outcomes based on patient-, surgery- and prosthesis-specific factors [167-169]. An important part of this study was the analysis conducted to examine the impact of potential confounders which are known to influence implant survival. This analysis sought to identify patient and surgeon factors, as well as prosthesis-related variables that may have contributed to the observed higher rate of revisions. This research studied a number of surgeons (after applying the exclusion criteria) to determine whether the poor results of outlier prostheses were influenced by one or a few surgeon(s). However, a more detailed analysis on the experience of surgeons and their surgical load (e.g. number of replacements performed per year) was not conducted due to some major limitations, such as the complexity of converting such information into classified meaningful inputs. This is a limitation in this study that needs further investigation in future work.

This research identified hip and knee outlier prostheses, taking into consideration all the known confounding factors such as the number of surgeons and the subset designs. A particular surgeon or a specific design of device could be responsible for a higher-than-anticipated rate of revision (HTARR). These two factors were not included in the model training but were studied through a careful post-analysis of the surgeon effects and subset details. This study also identified the surgeons with the most contribution to the performance of identified total hip and knee prostheses. The outcomes of the outlier prostheses by these surgeons are shown in Table 8.1.

Table 8. 1. The most significant interactions between surgeon *ID*s and prostheses.

Primary Total Conventional Knee for OA (1st January 2015 to 31st December 2019)					
Surgeon *ID*		**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs.Years (95% CI)**
1177	Device *IV*	15	197	649.9	2.31 (1.29, 3.81)
	Overall	21	219	721.5	2.91 (1.80, 4.45)
	Ex. outliers	6	22	71.6	8.38 (3.07, 18.24)
1218	Device *IV*	7	439	414.4	1.69 (0.68, 3.48)
	Overall	39	921	1963.8	1.98 (1.41, 2.71)
	Ex. outliers	32	482	1549.4	2.06 (1.41, 2.91)
1745	Device *IV*	12	106	374.5	3.2 (1.65, 5.6)
	Overall	15	231	561.7	2.67 (1.49, 4.40)
	Ex. outliers	3	125	187.2	1.60 (0.33, 4.68)
Primary Total Conventional Hip for OA (1st January 2015 to 31st December 2019)					
Surgeon *ID*		**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs.Years (95% CI)**
1745	Device *I*	9	113	144.5	6.23 (2.85, 11.82)
	Overall	10	145	258.4	3.87 (1.85, 7.12)
	Ex. outliers	1	28	95.6	1.04 (0.03, 5.83)
1177	Device *VI*	6	134	275.7	2.17 (0.8, 4.74)
	Overall	30	425	1303.9	2.30 (1.55, 3.28)
	Ex. outliers	24	291	1028.2	2.72 (1.49, 3.47)
1218	Device *VI*	16	366	404.2	3.96 (2.26, 6.43)
	Overall	43	763	1666.6	2.58 (1.87, 3.47)
	Ex. outliers	27	397	1262.4	2.14 (1.41, 3.11)
1745	Device *V*	9	112	144.4	6.23 (3.32, 12.74)
	Overall	10	145	258.4	3.87 (1.85, 7.12)
	Ex. outliers	1	28	95.6	1.04 (0.03, 5.83)

8.2 Primary Total Hip Comparator

Given the advancements in the design and use of hip prostheses, this research aimed first to develop a specific comparator for the early assessment of total hip prostheses. Currently, the standard hip prosthesis is constructed with modern bearing surfaces. Hence, all non-modern bearing surfaces were excluded in order to identify a modern comparator. Then devices with modular neck-stem design or those used for specific purposes including constrained, dual mobility, and head size smaller than 28 mm were excluded. Lastly, all remaining prostheses previously identified as having HTARR were also excluded. These exclusions progressively reduced the revision rate of the comparator. Therefore, contemporary device components can be a more effective and relevant comparator with greater sensitivity for the early assessment of newly-introduced prostheses.

The AOANJRR standardised approach detected additional femoral stems and acetabular components using the final modified comparator. The registry takes into account the complexities with small sample sizes, the effect of a single surgeon, and the confounding impact of associated device components [164, 165]. Therefore, these devices were investigated further to determine the effect of surgeons and subset designs. Such identification usually has a significant positive influence on the clinical outcomes of patients. However, a careful comparative analysis needs to be undertaken to monitor and record the performance of joint arthroplasty prostheses. Surgeons, regulatory authorities, and industry should be aware of individual hip prostheses identified through a practical and modern initial screening test.

8.3 Primary Total Knee Comparator

There are variations in primary total knee outcomes by stability design referring to certain prosthetic structures designed to substitute for the inherent stability of knee ligaments. Most of the implanted prostheses are established cruciate retaining (CR) or posterior stabilised (PS) prostheses with long-term follow-ups [4]. These two designs remained the most popular and were used in 90.8% of all the primary procedures [34, 40, 66]. The current knee comparator does not differentiate amongst other total complex or conventional procedures. This may cause the detection of less conventional and more complex design devices within the community. Our results showed that the final conventional comparator, which includes only the conventionally-designed prostheses, has a lower revision rate than the current comparator. A comparator of complex prostheses was also identified with a higher revision rate than the current comparator to better reflect high-risk devices used for specific purposes.

The use of the modified comparator groups resulted in more conventional and less complex prostheses detected by the AOANJRR standardised approach. Although the comparator was modified in terms of the stability design, there was no further subdivision by patella, fixation, mobility, or technology assist (including robotic assistance, image-derived instrumentation, or computer navigation). Further subdivisions may provide meaningless comparator groups with too-small sample sizes at the time of writing this study. However, the registry is aware that other factors can be further investigated for more relevant and specific comparator groups. For example, using only the total knee prostheses with patellar resurfacing components may improve the sensitivity of the conventional comparator [5, 34, 40]. Regular reconsiderations of initial screening are needed as modifications occur in the use and design of total knee prostheses.

8.4 Methods for Outlier Detection

Many diverse models exist for analysing and studying high-dimensional survival data. While some statistical models and implementations contain tuning automatism of a parameter internally, others may necessitate the user to modify defaults accurately. Survival analysis agreements with the analysis of failure times or time to events are observed in various application fields, such as medical statistics. The main interest of researchers usually is in computing the effect of other variables on the survival time. For example, the Cox proportional hazards model is a standard method used to analyse the impact of clinical variables on the outcome.

An established research method is the development of appropriate multivariate survival models in addition to the comparative study of their potential to generalize using unseen data. However, new challenges have appeared in recent years due to the quick expansion of data collection technologies and computer science. One particular challenge involves the reliable and simultaneous measurement of thousands of variables such as patient data and prosthesis attributes. Common techniques comprise linear regression methods that penalize extreme parameter estimates with a shrinkage format such as a Ridge penalty [213], a Lasso penalty [214] or a combined version called elastic net [215]. The last two have the advantage of automatically implementing feature selection for subsequent interpretation. However, it has been found that Ridge regression – which keeps all parameters in its final coefficient vector –often performs superior with a greater power of prediction [216, 217].

Although comprehensive comparisons have been performed between ML statistical methods, comparative study of survival methods is still commonly conducted. There is a strong need for reproducible and objective comparative studies in the field of joint arthroplasty surgeries. Despite the aforementioned penalized approaches, survival trees/forests seem promising approaches that have not been severely compared in the high-dimensional survival setting. The effectiveness of the developed models may vary by tuning of hyper-parameters, as the default settings are often not optimal. Even experts might struggle with the tuning process but prefiltering has been recommended and proven to be effective. For example, most variables are unrelated to survival outcomes in typical signal detection and can be regarded as irrelevant or even detrimental to prediction.

Various filtering methods can be used, requiring the analyst to set the respective control parameters. The in-depth evaluation of all algorithm and parameter combinations is impractical even if parallelisation is employed, especially for the high-dimensional structures. A modern approach to solving this dilemma is to make all choices through an efficient black-box optimisation that considers the desired performance measure. Modern methods are exclusively customised for the characteristics of optimisation problems. The emerging research field has become known as a hot modern topic called algorithm configuration. Two leading procedures are iterated racing [218] and model-based optimization [219].

8.5 Survival Analysis

In survival analysis, the time to a particular event (first revision) is observed. The predefined event may not be observable due to the censoring for a patient who may die or survive until the end of this study. This important concept is called right-censoring and holds essential information about the event that did not occur before a certain point. Therefore, one primary goal is to involve both non-censored and censored observations. This research assumed that there are *n* patients with a survival time *"ti"* for each of them. These survival times are supposed to be generated from a non-negative random variable *T* with cumulative distribution function *F(t) and* density *f(t).* The probability of surviving more than *t* is specified by the survival function *S(t)*:

$$S(t) = P(\{T > t\}) = \int_t^{\infty} f(u)\, du = 1 - F(t)$$

The number of observations is defined as *ni* and also, when it is still under risk at *ti* is defined as the number of risk at *ti*. The Kaplan–Meier estimator can estimate hazard function λ(t) that is closely associated with the survival function and express the risk for an event (revision) at a definite point in time *t* with respect to surviving until *t* and the cumulative hazard function *Λ(t)*:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{S(t) - S(t + \Delta t)}{\Delta t\, S(t)} = \frac{f(t)}{S(t)}, \Lambda(t) = \int_0^t \lambda(u)\, du = -\log S(t)$$

This study assumed the censoring to be non-informative, which means that the censoring distribution of data contains no information about the surveillance. In addition, a covariate vector $x_i \in Rp$ was observed for each patient, which then formed the *(n × p)*-matrix by all covariate vectors. The covariates can only include a limited number of variables (e.g., age and gender) or constantly a large number of orthopaedic variables. They are mostly considered to predict the risk of occurring an event expressed by $\lambda(t)$ or $\Lambda(t)$. The concordance index (C-index) is a standard performance measure to assess survival models [220]. The C-index, including a correction for ties, is estimated by maximizing the partial log-likelihood.

8.6 Feature Selection and Algorithms

As a strategy for mining high-dimensional datasets, feature selection shows this ability to be efficient and effective. The aim of feature selection is to build more straightforward and understandable models and improve data-mining performance. The current proliferation of big data is suggested considerable challenges and research opportunities. It is required to apply data mining strategies to discover knowledge from big data. Learning models tend to overfit with many features, leading to performance decay on unseen data. High-dimensional data can considerably increase the memory storage necessities and computational requirements.

Real-world data hold a large number of unrelated, redundant, and noisy variables that can be handled by one of the most influential tools called dimensionality reduction. On the one hand, feature extraction changes the dataset with original high-dimensional variables to a new space with lower dimensionality. The newly-created variable space can typically be a nonlinear or linear combination of the original variables. On the other hand, feature selection directly chooses a subgroup of relevant variables for model development [195, 221].

Basic filtering methods exist to handle the correlations and there are more elaborate alternates that demand extensive computational resources. The pre-selection of features is a two-edged sword: the model may be untrustworthy as it might not reveal potentially useful information. By contrast, limiting the features to a logical subset can meaningfully increase the performance of models and preliminary dimension reduction decreases the computational necessities [21]. A backward selection procedure may help provide a more informative subset of variables.

The removal of noisy variables also decreases computational cost and simultaneously prevents significant information loss or degradation of learning performance. Both feature selection and extraction can improve learning performance, reduce memory storage, and fit models with better generalization. Feature selection keeps the physical meanings of the original variables and provides models with greater interpretability, and is frequently desired in many applications such as medical analytics.

Algorithms have different unique implementations. Survival trees use binary selections to split data into sets of similar hazard rates recursively. A rule for binary splits is to search all variables and their corresponding probable cut points to minimize the *P*-value of the log-rank test [222]. For prediction purposes, new data are released to the tree and a cumulative hazard function $\Lambda(t)$ is created for each separate terminal node. This function can be used to compute the risk of revision with respect to the follow-up time. Random survival forests are built by fitting survival trees on bootstrapped data samples and randomly sampling candidate feature sets for each node.

In the case of Cox regression, Lambda represented the complexity of the model determined via 10-fold Cross-Validation (CV). An acceptable value for lambda is typically identified using CV to provide a stable model with the minimum variance in the reported outcomes. This occurs by a random subdivision of the dataset into ten parts, and each block then acts as the test set per CV iteration. Cross-validation is a common resampling process for the assessment and comparative evaluation of the model [223]. A rather high number of folds was selected as there were enough samples to evaluate the C-index in a meaningful way. Note that the proposed models were all developed with a tuning of the most critical parameters using CV.

8.7 Machine Learning Approach for the Identification of Outlier Prostheses

The AOANJRR multi-stage approach ignores the order of time. This means that although the registry is successful at evaluating the performance of recently-introduced devices, those with delayed onset of higher revision rates are not simply detected by the same method. It also does not limit the confounding effect of the potential factors. Ideally, a survival method to identify outliers should determine the confounding effects of patient characteristics and other associated components using a time-to-first revision outcome. Stage *I* of the standardised approach does not account appropriately for the changes in the rate of revision over time. This constraint may cause difficulty in detecting the outlier prostheses later in the follow-up period [173]. In addition, stage *II* is a further analysis only on a limited number of confounding variables such as age, gender and primary diagnosis.

In medical research, predictive models are often used to derive patterns from diagnoses and treatments. Examples include data-driven predictions of drug effects, the detection of comorbidity groups in autism spectrum disorders, and the identification of type 2 diabetes subgroups [97, 98]. Machine learning (ML) seems promising for solving complex problems as many variables can be considered simultaneously and learning patterns are produced from empirical data. However, more investigations are needed to document the ability of feature selection in orthopaedics and the monitoring of joint arthroplasty, particularly total joint replacement. For example, the success of outlier detection relies on detecting appropriate component characteristics, and the process is compromised in the absence of relevant attributes. The contrary may be a concern because when too many features are used to describe components, outliers may be difficult to detect. One possibility to address this issue is to use data from several joint registries with information about the components of the same prostheses.

In medical sciences, supervised ML is widely used to train models with a known set of predictors and outcomes. The supervised algorithms are able to identify the predictors most strongly associated with the outcome. Larger sample sizes may improve the performance of ML in the variable selection or prediction problems. The potential to handle big data with high-correlated structures may assist clinicians by providing information about the components of the medical device. In addition, retrieval studies add significant insights on the mechanisms of failed implants, and should be used in conjunction with joint replacement registry reports [224, 225]. Implant retrieval studies and joint registry data analysis of THR and TKR are multidisciplinary areas that require contributions from clinicians, engineers and data scientists. To date, multivariable predictive models have been developed for THR and TKR using patient-reported factors and image-based data. Perhaps now is the right time to enter a new era of THR and TKR by developing decision-making support systems.

The principal objective of evaluating ML techniques was to investigate their potential in monitoring the performance of joint arthroplasty components. There was variation in the outcome of methods employed to detect hip and knee outliers with respect to the standardised approach. Hip models were trained with smaller sample sizes and more highly-correlated inputs than the knee models due to entering the individual components rather than combinations. By contrast, more covariates were defined in the case of total knee replacements in order to consider further complex interactions. It is noted that the ML approach necessitates clinical knowledge prior to selecting a reasonable number of input factors.

More feature selection techniques could be evaluated using shared data from joint registries to improve signal detection efforts. Machine learning has limitations in identifying outlier prostheses with concerns about accurately interpreting statistics to indicate the impact of variables on the outcome [226, 227]. Random forests are unable to specify variable effects in a substantively meaningful approach, and it is also more challenging to achieve a substantive understanding of variable effects [6].

The AOANJRR standard provided an acceptable criterion by which to assess the effectiveness of the methods on both reported and non-reported prostheses. We suggest that ML can be used as a supplementary approach in the outlier detection and tree-based methods offer better performance for data mining [175, 227]. To date, several techniques (e.g., permutation importance) have been used to interpret variable effects determined by random survival modelling [227]. Permutation importance does not characterize variable effects, and it only describes a conditional strength.

This study found that the estimate of variable importance is less biased and more accurate when the sample size is increased with a corresponding increase in the computational time. Both of the proposed ML methods reduced the dimensionality of our complex data to a subset of more informative inputs. This was done by excluding noisy variables that could possibly improve the prediction error rate to select the most significant predictors. Hyperparameters were also tuned carefully to achieve the most informative predictors. Random forests require the user to examine the impact of tuned hyperparameters on the error rate. A more computational cost should be expected for RSF than for the parametric or semi-parametric models.

The conventional parametric models such as multiple linear regression necessitate a correct model specification. To the extent that the fitted Cox models rely on a more conventional approach of measuring variable significance, it may be a more desirable option for similar feature selection problems. However, limitations exist for the Cox modelling regarding the extrapolation and model misspecification [228]. Other advantages of the Cox models include higher efficiency in computational time and reporting the adjusted predictor's strength when there is a need to maintain the false discovery rate (FDR) at 0.05. Conversely, calculating a sufficient number of permutations with respect to the FDR rate was computationally too burdensome for the RSF approach.

In future applications, Cox and random survival may support the initial screening test to effectively monitor the performance of prostheses. A single model is not necessarily the best as the inclusion or exclusion of predictors can affect the strength and sign of predictors. Some points of the two methods are sorted in order. First, more components were detected using RSF within the total conventional hip community. Second, the primary motivation of this study was to limit both device- and patient-level confounding when identifying harmful prostheses, and the two approaches could handle this difficulty. The Cox was conditioned on device and patient characteristics simultaneously while the RSF had many competing variables for splitting.

Overall, the RSF backward selection was more aligned with the AOANJRR standard as random subsets of variables were used per node to grow trees. In addition, the correlated variables were selected independently, leading to the interruption of multicollinearity [188]. This feature of the RSF provides an attractive approach for explorative variable selection; however, false-positive discoveries due to overfitting are still considered a major problem [189]. Machine learning identified additional hip prostheses to the AOANJRR standard, and some of those were newly detected for the first time based on the registered historical data. Future studies can apply the developed approaches to monitor other arthroplasty devices such as those for shoulders. Utilizing prediction to understand the link between inputs and the outcome may improve decision makings for the early identification of outlier prostheses.

Chapter 9. Conclusions

9.1 Conclusion remarks

One of the main functions of all joint registries (JRs) is to enable the comparison of the performance of individuals, prosthesis combinations, or an entire class of devices. This study illustrates that increasing the relevance of the comparator could be helpful for the early identification of a higher number of outlier prostheses. In addition, the most challenging part was the reduction of the confounding effects by taking into account the patient factors and prosthesis attributes. For example, several total knee replacements (TKRs) have a higher revision rate only with posterior stabilized or cementless versions, and total hip replacements (THRs) show the same risk with smaller head sizes. The results enable surgeons to make an informed choice of devices, and are essential for registries to identify variations, leading to the adoption of best practices.

Several prostheses currently being used extensively in the Australian market were identified for the first time. The use of modified comparator groups led to identifying fewer complex knee designs and additional conventional hip and knee prostheses that pose a risk. The final comparator groups had an estimated 10-year CPR of 4.3% (4.2, 4.41) for conventional THR, 5.2% (5.1, 5.3) for conventional TKR, and 10.3% (8.6, 12.0) for Complex TKR. The publication of these data could contribute to better clinical outcomes by reducing the revision rate. This research studied in detail and statistically reported the performance of prostheses utilising the modified comparator. This includes patient factors, prosthesis attributes, and the potential interaction between the confounding factors.

This study evaluated random survival forest (RSF) and Cox regression based on their ability to detect the outliers identified by the standardised approach. In addition, the performance of these hip and knee devices was assessed with a view to limiting the effect of potential confounding factors. This study showed that the random survival approach was more comparable to the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) standard in terms of detecting more outlier prostheses. However, the Cox regression has a significant advantage in terms of computational cost, interpreting variable effects, and documenting confounding effects. Overall, five prostheses met all the criteria established by the approaches (P value < .05) and several devices had higher than expected revision rates.

The registries have the common goal of improving joint arthroplasty outcomes over time. The early identification of outliers may be associated with a substantial reduction in usage of a certain prosthesis and the subsequent withdrawal of the device from the Australian market. The registry and NICE (National Institute for Health and Care Excellence) currently recommend the modern comparator for the early detection of primary total hip and knee prostheses. In addition, random survival and Cox regression techniques might offer a supplementary approach to improve the statistical process strengthened by the registry. The AOANJRR verifies this research findings that sensitively improved the early assessment of prostheses.

A number of total knee prostheses were identified as combinations but more consideration should be given to reporting the individual components. Monitoring the real performance of prostheses with more combinations can better reflect the performance of individual components. Further weight is also given to the argument that the non-modern bearing surfaces were an issue with a broad range of joint arthroplasty prostheses. To the best of our knowledge, this is the first comparative study to report the poorer performance of non-modern THR compared to the modern bearing surfaces. The continued use of such prostheses may increase the risk of using poor-performing prostheses for more patients with osteoarthritis.

The AOANJRR aims to develop a better understanding of confounding factors associated with outlier detection. According to the machine learning chapters, the most significant device-related covariates include head size for the initial screening of hip devices, and the stability and fixation for that of knee prostheses. Moreover, the conventional statistical analyses showed variations in the outcome of identified total hip prostheses with the femoral head size, bearing surface and the associated device component, and knee outliers in regard to bearing surfaces, method of fixation and the patella usage.

This finding suggests the importance of identifying the confounding factors and evaluating their impacts on the detection. Machine learning seems promising as an initial screening method for a more effective assessment of prostheses. Utilizing prediction to understand the variables linked with the outcome may improve shared decision-making, leading to fewer patients at risk of receiving poor devices. The outcome may cause a considerable reduction in the number of patients exposed to the outlier prostheses.

9.2 Suggestions for future works

The AOANJRR has been instrumental in improving the joint arthroplasty outcomes, and this research on the registry data could have an increasingly widespread global influence. The registry has worked closely with all stakeholders involved in THRs and TKRs including industry, surgeons, hospitals, government and regulatory bodies, medical insurers and patients. Collaboration between JRs for the purpose of sharing data, will enable researchers to conduct for more extensive analyses of the prosthesis outcomes of surgery performed internationally. Future studies can apply the proposed method to various classes of device components used for arthroplasty surgeries. The concept of prediction models to understand the significance of variables may have considerable potential to provide important context for the initial screening of prosthetic devices. In addition, this research included significant clinically-known attributes but other factors related to surgeons and subset designs can be investigated.

Reference list

1. Rory J Ferguson, Antony JR Palmer, Adrian Taylor, Martyn L Porter, Henrik Malchau, and Sion Glyn-Jones, *Hip replacement.* The Lancet, 2018. **392**(10158): p. 1662-1671.
2. Philip Wiles, *The surgery of the osteo-arthritic hip.* British Journal of Surgery, 1958. **45**(193): p. 488-497.
3. Ian D Learmonth, Claire Young, and Cecil Rorabeck, *The operation of the century: total hip replacement.* The Lancet, 2007. **370**(9597): p. 1508-1519.
4. Andrew J Price, Abtin Alvand, Anders Troelsen, Jeffrey N Katz, Gary Hooper, Alastair Gray, Andrew Carr, and David Beard, *Knee replacement.* The Lancet, 2018. **392**(10158): p. 1672-1682.
5. *Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). Hip, knee & shoulder arthroplasty: 2020 annual report. Adelaide: AOA, 2020.*
6. Sion Glyn-Jones, AJR Palmer, R Agricola, AJ Price, TL Vincent, H Weinans, and AJ Carr, *Osteoarthritis.* The Lancet, 2015. **386**(9991): p. 376-387.
7. Elena Losina, Thomas S Thornhill, Benjamin N Rome, John Wright, and Jeffrey N Katz, *The dramatic increase in total knee replacement utilization rates in the United States cannot be fully explained by growth in population size and the obesity epidemic.* The Journal of Bone and Joint Surgery. American volume., 2012. **94**(3): p. 201.
8. Amresh D Hanchate, Alok Kapoor, Jeffrey N Katz, Danny McCormick, Karen E Lasser, Chen Feng, Meredith G Manze, and Nancy R Kressin, *Massachusetts health reform and disparities in joint replacement use: difference in differences study.* bmj, 2015. **350**.
9. Richard N de Steiger, Jacqueline R Hang, Lisa N Miller, Stephen E Graves, and David C %J JBJS Davidson, *Five-year results of the ASR XL acetabular system and the ASR hip resurfacing system: an analysis from the Australian Orthopaedic Association National Joint Replacement Registry.* 2011. **93**(24): p. 2287-2293.
10. *Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR). Hip, knee & shoulder arthroplasty. Adelaide: AOA.* 2012.
11. Jill Dawson, Ray Fitzpatrick, Andrew Carr, and David Murray, *Questionnaire on the perceptions of patients about total hip replacement.* The Journal of bone and joint surgery. British volume, 1996. **78**(2): p. 185-190.
12. Nancy J Devlin and Richard Brooks, *EQ-5D and the EuroQol group: past, present and future.* Applied health economics and health policy, 2017. **15**(2): p. 127-137.
13. Kristina Harris, Jill Dawson, Elizabeth Gibbons, Chris R Lim, David J Beard, Raymond Fitzpatrick, and Andrew J Price, *Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty.* Patient related outcome measures, 2016. **7**: p. 101.
14. AT Bremner-Smith, P Ewings, and AE Weale, *Knee scores in a 'normal'elderly population.* The Knee, 2004. **11**(4): p. 279-282.
15. Carlos J Lavernia, Victor Hugo Hernandez, and Mark D Rossi, *Payment analysis of total hip replacement.* Current Opinion in Orthopaedics, 2007. **18**(1): p. 23-27.
16. Matthias Vogl, Rainer Wilkesmann, Christian Lausmann, and Werner Plötz, *The impact of preoperative patient characteristics on the cost-effectiveness of total hip replacement: a cohort study.* BMC health services research, 2014. **14**(1): p. 1-12.
17. Meghan E Daigle, Alexander M Weinstein, Jeffrey N Katz, and Elena Losina, *The cost-effectiveness of total joint arthroplasty: a systematic review of published literature.* Best practice & research Clinical rheumatology, 2012. **26**(5): p. 649-658.
18. Care Excellence, *Guide to the Methods of Technology Appraisal 2013 [Internet].* 2013.
19. Gillian A Hawker, Elizabeth M Badley, Ruth Croxford, Peter C Coyte, Richard H Glazier, Jun Guan, Bart J Harvey, JI Williams, and James G Wright, *A population-based nested case-control study of the costs of hip and knee replacement surgery.* Medical care, 2009: p. 732-741.
20. Felipe Eltit, Qiong Wang, Rizhi %J Frontiers in bioengineering Wang, and biotechnology, *Mechanisms of adverse local tissue reactions to hip implants.* 2019. **7**.

21. S Rymaruk, A Razak, and R %J Journal of surgical case reports McGivney, *Metallosis, psoas abscess and infected hip prosthesis in a patient with bilateral metal on metal total hip replacement.* 2012. **2012**(5): p. 11-11.
22. Andrew Gratton, Bert Buford, Tarun Goswami, D Gaddy Kurten, and Larry %J Journal of the Mechanical Behavior of Materials Suva, *Failure modes of biomedical implants.* 2002. **13**(5-6): p. 297-314.
23. Donald W Howie and Margaret A McGee, *Wear and osteolysis in relation to prostheses design and materials*, in *Medical Applications of Titanium and Its Alloys: The Material and Biological Issues*. 1996, ASTM International.
24. Ian C Clarke, Pat Campbell, and Nir Kossovsky, *Debris-mediated osteolysis—a cascade phenomenon involving motion, wear, particulates, macrophage induction, and bone lysis*, in *Particulate debris from medical implants: mechanisms of formation and biological consequences*. 1992, ASTM International.
25. Yan Zhou, Ming-xue Shen, Zhen-bing Cai, Jin-fang Peng, and Min-hao %J Wear Zhu, *Study on dual rotary fretting wear behavior of Ti6Al4V titanium alloy.* 2017. **376**: p. 670-679.
26. Ying-Ying J Kao, Chelsea N Koch, Timothy M Wright, and Douglas E %J The Journal of arthroplasty Padgett, *Flexural rigidity, taper angle, and contact length affect fretting of the femoral stem trunnion in total hip arthroplasty.* 2016. **31**(9): p. 254-258.
27. Abimbola Oladokun, M Pettersson, M Bryant, Håkan Engqvist, Cecilia Persson, Richard Hall, A %J Tribology-Materials Neville, Surfaces, and Interfaces, *Fretting of CoCrMo and Ti6Al4V alloys in modular prostheses.* 2015. **9**(4): p. 165-173.
28. Jens Dargel, Johannes Oppermann, Gert-Peter Brüggemann, and Peer Eysel, *Dislocation following total hip replacement.* Deutsches Ärzteblatt International, 2014. **111**(51-52): p. 884.
29. Viktor Lindgren, Max Gordon, Per Wretenberg, Johan Kärrholm, and Göran Garellick, *Deep infection after total hip replacement: a method for national incidence surveillance.* Infection Control & Hospital Epidemiology, 2014. **35**(12): p. 1491-1496.
30. Herbert O Gbejuade, Andrew M Lovering, and Jason C Webb, *The role of microbial biofilms in prosthetic joint infections: a review.* Acta orthopaedica, 2015. **86**(2): p. 147-158.
31. Kevin J Mulhall, Hassan M Ghomrawi, Sean Scully, John J Callaghan, and Khaled J Saleh, *Current etiologies and modes of failure in total knee arthroplasty revision.* Clinical Orthopaedics and Related Research®, 2006. **446**: p. 45-50.
32. John J Callaghan, Michael R O'Rourke, and Khaled J Saleh, *Why knees fail: lessons learned.* The Journal of arthroplasty, 2004. **19**(4): p. 31-34.
33. Michael B Vessely, Andrew L Whaley, W Scott Harmsen, Cathy D Schleck, and Daniel J Berry, *The Chitranjan Ranawat Award: Long-term survivorship and failure modes of 1000 cemented condylar total knee arthroplasties.* Clinical Orthopaedics and Related Research®, 2006. **452**: p. 28-34.
34. *Swedish Knee Arthroplasty Register. Swedish Knee Arthroplasty Register Annual report* 2020.
35. Douglas DR Naudie, Deborah J Ammeen, Gerard A Engh, and Cecil H Rorabeck, *Wear and osteolysis around total knee arthroplasty.* JAAOS-Journal of the American Academy of Orthopaedic Surgeons, 2007. **15**(1): p. 53-64.
36. VI Roberts, CNA Esler, and WM Harper, *A 15-year follow-up study of 4606 primary total knee replacements.* The Journal of bone and joint surgery. British volume, 2007. **89**(11): p. 1452-1456.
37. *Australian Orthopaedic Association National Joint Replacement Registry. Hip, knee & shoulder arthroplasty: annual report 2007.*
38. Steven M Kurtz, Kevin L Ong, Edward Lau, Kevin J Bozic, Daniel Berry, and Javad Parvizi, *Prosthetic joint infection risk after TKA in the Medicare population.* Clinical Orthopaedics and Related Research®, 2010. **468**(1): p. 52-56.
39. E Anthony Rankin, Graciela S Alarcon, Rowland W Chang, and Leo M Cooney Jr, *NIH Consensus Statement on total knee replacement December 8-10, 2003.* Journal of Bone and Joint Surgery, 2004. **86**(6): p. 1328.
40. *New Zealand Joint Registry. The New Zealand Joint Registry: twenty-one year report January 1999 to December 2019. 2020.*

41. Patrick O'Toole, Mitchell G Maltenfort, Antonia F Chen, and Javad Parvizi, *Projected increase in periprosthetic joint infections secondary to rise in diabetes and obesity.* The Journal of arthroplasty, 2016. **31**(1): p. 7-10.
42. Benjamin Zmistowski, Craig Della Valle, Thomas W Bauer, Konstantinos N Malizos, Abbas Alavi, Hani Bedair, Robert E Booth, Peter Choong, Carl Deirmengian, and Garth D Ehrlich, *Diagnosis of periprosthetic joint infection.* The Journal of arthroplasty, 2014. **29**(2): p. 77-83.
43. FS Haddad and DA George, *Can National Joint Registries play a role in improving our understanding of periprosthetic infections?* 2016.
44. *16th Annual Report National Joint Registry for England, Wales, Northern Ireland and the Isle of Man - Annual Report 2019.*
45. Cody C Wyles, Jose H Jimenez-Almonte, Mohammad H Murad, German A Norambuena-Morales, Miguel E Cabanela, Rafael J Sierra, and Robert T Trousdale, *There are no differences in short-to mid-term survivorship among total hip-bearing surface options: a network meta-analysis.* Clinical Orthopaedics and Related Research®, 2015. **473**(6): p. 2031-2041.
46. Robert B Bourne, Bert M Chesworth, Aileen M Davis, Nizar N Mahomed, and Kory DJ Charron, *Patient satisfaction after total knee arthroplasty: who is satisfied and who is not?* Clinical Orthopaedics and Related Research®, 2010. **468**(1): p. 57-63.
47. *National Joint Registry for England, Wales, Northern Ireland and the Isle of Man.* 16th annual report, 2019.
48. BM Wroblewski, PD Siney, and PA Fleming, *Charnley low-friction arthroplasty: survival patterns to 38 years.* The Journal of bone and joint surgery. British volume, 2007. **89**(8): p. 1015-1018.
49. R Kallala, P Anderson, S Morris, and FS Haddad, *The cost analysis of cemented versus cementless total hip replacement operations on the NHS.* The bone & joint journal, 2013. **95**(7): p. 874-876.
50. GJ Hooper, AG Rothwell, M Stringer, and C Frampton, *Revision following cemented and uncemented primary total hip replacement: a seven-year analysis from the New Zealand Joint Registry.* The Journal of bone and joint surgery. British volume, 2009. **91**(4): p. 451-458.
51. Daniel J Berry, W Scott Harmsen, Miguel E Cabanela, and Bernard F Morrey, *Twenty-five-year survivorship of two thousand consecutive primary Charnley total hip replacements: factors affecting survivorship of acetabular and femoral components.* JBJS, 2002. **84**(2): p. 171-177.
52. Donald W Howie, Oksana T Holubowycz, Robert Middleton, and Large Articulation Study Group, *Large femoral heads decrease the incidence of dislocation after total hip arthroplasty: a randomized controlled trial.* JBJS, 2012. **94**(12): p. 1095-1102.
53. Donald W Howie, Oksana T Holubowycz, and Stuart A Callary, *The wear rate of highly cross-linked polyethylene in total hip replacement is not increased by large articulations: a randomized controlled trial.* JBJS, 2016. **98**(21): p. 1786-1793.
54. Brigitte M Jolles and Earl R Bogoch, *Posterior versus lateral surgical approach for total hip arthroplasty in adults with osteoarthritis.* Cochrane Database of Systematic Reviews, 2003(4).
55. JV Lindgren, Per Wretenberg, J Kärrholm, G Garellick, and O Rolfson, *Patient-reported outcome is influenced by surgical approach in total hip replacement: a study of the Swedish Hip Arthroplasty Register including 42 233 patients.* The bone & joint journal, 2014. **96**(5): p. 590-596.
56. G Meermans, S Konan, R Das, A Volpin, and FS Haddad, *The direct anterior approach in total hip arthroplasty: a systematic review of the literature.* The bone & joint journal, 2017. **99**(6): p. 732-740.
57. Joseph D Maratt, Joel J Gagnier, Paul D Butler, Brian R Hallstrom, Andrew G Urquhart, and Karl C Roberts, *No difference in dislocation seen in anterior vs posterior approach total hip arthroplasty.* The Journal of arthroplasty, 2016. **31**(9): p. 127-130.
58. James Chow, Brad Penenberg, and Stephen Murphy, *Modified micro-superior percutaneously-assisted total hip: early experiences & case reports.* Current reviews in musculoskeletal medicine, 2011. **4**(3): p. 146.
59. G Grammatopoulos, GER Thomas, H Pandit, DJ Beard, HS Gill, and DW Murray, *The effect of orientation of the acetabular component on outcome following total hip arthroplasty with small diameter hard-on-soft bearings.* The bone & joint journal, 2015. **97**(2): p. 164-172.

60. Yu-Lin Li, Jun Jia, Qiang Wu, Guang-Zhi Ning, Qiu-Li Wu, and Shi-Qing Feng, *Evidence-based computer-navigated total hip arthroplasty: an updated analysis of randomized controlled trials.* European Journal of Orthopaedic Surgery & Traumatology, 2014. **24**(4): p. 531-538.
61. Richard N De Steiger, Yen-Liang Liu, and Stephen E Graves, *Computer navigation for total knee arthroplasty reduces revision rate for patients less than sixty-five years of age.* JBJS, 2015. **97**(8): p. 635-642.
62. Benjamin G Domb, Youssef F El Bitar, Adam Y Sadik, Christine E Stake, and Itamar B Botser, *Comparison of robotic-assisted and conventional acetabular cup placement in THA: a matched-pair controlled study.* Clinical Orthopaedics and Related Research®, 2014. **472**(1): p. 329-336.
63. Illgen RL Nd, Brandon R Bukowski, Rasheed Abiola, Paul Anderson, Morad Chughtai, Anton Khlopas, and Michael A Mont, *Robotic-assisted total hip arthroplasty: outcomes at minimum two-year follow-up.* Surgical technology international, 2017. **30**: p. 365-372.
64. L Spencer-Gardner, J Pierrepont, M Topham, J Baré, S McMahon, and AJ Shimmin, *Patient-specific instrumentation improves the accuracy of acetabular component placement in total hip arthroplasty.* The bone & joint journal, 2016. **98**(10): p. 1342-1346.
65. Andrew J Carr, Otto Robertsson, Stephen Graves, Andrew J Price, Nigel K Arden, Andrew Judge, and David J %J The Lancet Beard, *Knee replacement.* 2012. **379**(9823): p. 1331-1340.
66. *American Joint Replacement Registry. American Joint Registry 2020 annual report. 2020.*
67. Richard N de Steiger, Orhun Muratoglu, Michelle Lorimer, Alana R Cuthbert, and Stephen E Graves, *Lower prosthesis-specific 10-year revision rate with crosslinked than with non-crosslinked polyethylene in primary total knee arthroplasty: 386,104 procedures from the Australian Orthopaedic Association National Joint Replacement Registry.* Acta orthopaedica, 2015. **86**(6): p. 721-727.
68. Audrey K Nebergall, Anders Troelsen, Harry E Rubash, Henrik Malchau, Ola Rolfson, and Meridith E Greene, *Five-year experience of vitamin e–diffused highly cross-linked polyethylene wear in total hip arthroplasty assessed by radiostereometric analysis.* The Journal of arthroplasty, 2016. **31**(6): p. 1251-1255.
69. John N Insall, Roberto Binazzi, Michael Soudry, and Luiz A Mestriner, *Total knee arthroplasty.* Clinical orthopaedics and related research, 1985(192): p. 13-22.
70. C Rivière, F Iranpour, E Auvinet, S Howell, P-A Vendittoli, J Cobb, and S Parratte, *Alignment options for total knee arthroplasty: a systematic review.* Orthopaedics & Traumatology: Surgery & Research, 2017. **103**(7): p. 1047-1056.
71. Simon W Young, Matthew Walker, Ali Bayan, Toby Briant-Evans, Paul Pavlou, and Bill Farrington, *No difference in 2-year functional outcomes using kinematic versus mechanical alignment in TKA: a randomized controlled clinical trial.* Arthroscopy, 2017. **33**(10): p. e115.
72. P Maxwell Courtney and Gwo-Chin Lee, *Early outcomes of kinematic alignment in primary total knee arthroplasty: a meta-analysis of the literature.* The Journal of arthroplasty, 2017. **32**(6): p. 2028-2032. e1.
73. Edward Burn, Alexander D Liddle, Thomas W Hamilton, Andrew Judge, Hemant G Pandit, David W Murray, and Rafael Pinedo-Villanueva, *Cost-effectiveness of unicompartmental compared with total knee replacement: a population-based study using data from the National Joint Registry for England and Wales.* BMJ open, 2018. **8**(4): p. e020977.
74. Vikas Kulshrestha, Barun Datta, Santhosh Kumar, and Gaurav Mittal, *Outcome of unicondylar knee arthroplasty vs total knee arthroplasty for early medial compartment arthritis: a randomized study.* The Journal of arthroplasty, 2017. **32**(5): p. 1460-1469.
75. David Beard, Andrew Price, Jonathan Cook, Ray Fitzpatrick, Andrew Carr, Marion Campbell, Helen Doll, Helen Campbell, Nigel Arden, and Cushla Cooper, *Total or Partial Knee Arthroplasty Trial-TOPKAT: study protocol for a randomised controlled trial.* Trials, 2013. **14**(1): p. 1-11.
76. AD Liddle, H Pandit, A Judge, and DW Murray, *Optimal usage of unicompartmental knee arthroplasty: a study of 41 986 cases from the National Joint Registry for England and Wales.* The bone & joint journal, 2015. **97**(11): p. 1506-1511.
77. Kirill Gromov, Mounim Korchi, Morten G Thomsen, Henrik Husted, and Anders Troelsen, *What is the optimal alignment of the tibial and femoral components in knee arthroplasty? An overview of the literature.* Acta orthopaedica, 2014. **85**(5): p. 480-487.

78. Jelle P van der List, Harshvardhan Chawla, Leo Joskowicz, and Andrew D Pearle, *Current state of computer navigation and robotics in unicompartmental and total knee arthroplasty: a systematic review with meta-analysis.* Knee Surgery, Sports Traumatology, Arthroscopy, 2016. **24**(11): p. 3482-3495.
79. Otto Robertsson, *Knee arthroplasty registers.* The Journal of bone and joint surgery. British volume, 2007. **89**(1): p. 1-4.
80. Nancy A Dreyer and Sarah Garner, *Registries for robust evidence.* Jama, 2009. **302**(7): p. 790-791.
81. Stephen E Graves, *The value of arthroplasty registry data*. 2010, Taylor & Francis.
82. Alex Allepuz, Leif Havelin, Thomas Barber, Art Sedrakyan, Stephen Graves, Barbara Bordini, Daniel Hoeffel, Guy Cafri, and Elizabeth Paxton, *Effect of femoral head size on metal-on-HXLPE hip arthroplasty outcome in a combined analysis of six national and regional registries.* The Journal of Bone and Joint Surgery. American volume, 2014. **96**(Suppl 1): p. 12-18.
83. Ove Furnes, Elizabeth Paxton, Guy Cafri, Stephen Graves, Barbara Bordini, Thomas Comfort, Moises Coll Rivas, Samprit Banerjee, and Art Sedrakyan, *Distributed analysis of hip implants using six national and regional registries: comparing metal-on-metal with metal-on-highly cross-linked polyethylene bearings in cementless total hip arthroplasty in young patients.* The Journal of Bone and Joint Surgery. American volume, 2014. **96**(Suppl 1): p. 25-33.
84. Elizabeth Paxton, Guy Cafri, Leif Havelin, Susanna Stea, Francesc Pallisó, Stephen Graves, Daniel Hoeffel, and Art Sedrakyan, *Risk of revision following total hip arthroplasty: metal-on-conventional polyethylene compared with metal-on-highly cross-linked polyethylene bearing surfaces: international results from six registries.* JBJS, 2014. **96**(Suppl 1): p. 19-24.
85. Art Sedrakyan, Stephen Graves, Barbara Bordini, Miquel Pons, Leif Havelin, Susan Mehle, Elizabeth Paxton, Thomas Barber, and Guy Cafri, *Comparative effectiveness of ceramic-on-ceramic implants in stemmed hip replacement: a multinational study of six national and regional registries.* The Journal of bone and joint surgery. American volume, 2014. **96**(Suppl 1): p. 34-41.
86. Susanna Stea, Thomas Comfort, Art Sedrakyan, Leif Havelin, Marcella Marinelli, Thomas Barber, Elizabeth Paxton, Samprit Banerjee, Abby J Isaacs, and Stephen Graves, *Multinational comprehensive evaluation of the fixation method used in hip replacement: interaction with age in context.* JBJS, 2014. **96**(Suppl 1): p. 42-51.
87. Charles F Dillon, Elizabeth K Rasch, Qiuping Gu, and Rosemarie Hirsch, *Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991-94.* The Journal of rheumatology, 2006. **33**(11): p. 2271-2279.
88. Jaakko Julin, Esa Jämsen, Timo Puolakka, Yrjö T Konttinen, and Teemu Moilanen, *Younger age increases the risk of early prosthesis failure following primary total knee replacement for osteoarthritis: a follow-up study of 32,019 total knee replacements in the Finnish Arthroplasty Register.* Acta orthopaedica, 2010. **81**(4): p. 413-419.
89. Ove Furnes, Birgitte Espehaug, SA Lie, SE Vollset, LB Engesaeter, and LI Havelin, *Failure mechanisms after unicompartmental and tricompartmental primary knee replacement with cement.* JBJS, 2007. **89**(3): p. 519-525.
90. Todd Borus and Thomas Thornhill, *Unicompartmental knee arthroplasty.* JAAOS-Journal of the American Academy of Orthopaedic Surgeons, 2008. **16**(1): p. 9-18.
91. Laurie MC Faro and Rik Huiskes, *Quality assurance of joint replacement legal regulation and medical judgement.* Acta Orthopaedica Scandinavica, 1992. **63**(sup250): p. 1-33.
92. William L Healy and Richard Iorio, *Implant selection and cost for total joint arthroplasty: conflict between surgeons and hospitals.* Clinical Orthopaedics and Related Research®, 2007. **457**: p. 57-63.
93. Zosia Kmietowicz, *Patients who have new types of hip and knee replacement are more likely to need revision*. 2008, British Medical Journal Publishing Group.
94. *Swedish Hip Arthroplasty Register. The Swedish Hip Arthroplasty Register: annual report.* 2019.
95. Anna Floegel, Norbert Stefan, Zhonghao Yu, Kristin Mühlenbruch, Dagmar Drogan, Hans-Georg Joost, Andreas Fritsche, Hans-Ulrich Häring, Martin Hrabě de Angelis, and Annette Peters, *Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach.* Diabetes, 2013. **62**(2): p. 639-648.

96. SL Hardoon, JD Lewsey, PJ Gregg, BC Reeves, JHP %J The Journal of bone van der Meulen, and joint surgery. British volume, *Continuous monitoring of the performance of hip prostheses.* 2006. **88**(6): p. 716-720.
97. Finale Doshi-Velez, Yaorong Ge, and Isaac %J Pediatrics Kohane, *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis.* 2014. **133**(1): p. e54-e63.
98. Matthias B Schulze, Kurt Hoffmann, Heiner Boeing, Jakob Linseisen, Sabine Rohrmann, Matthias Möhlig, Andreas FH Pfeiffer, Joachim Spranger, Claus Thamer, and Hans-Ulrich Häring, *An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes.* Diabetes care, 2007. **30**(3): p. 510-515.
99. Federico Cabitza, Angela Locoro, and Giuseppe Banfi, *Machine learning in orthopedics: a literature review.* Frontiers in bioengineering and biotechnology, 2018. **6**: p. 75.
100. Nina Kozic, Stefan Weber, Philippe Büchler, Christian Lutz, Nils Reimers, Miguel Á González Ballester, and Mauricio Reyes, *Optimisation of orthopaedic implant design using statistical shape space analysis based on level sets.* Medical image analysis, 2010. **14**(3): p. 265-275.
101. Ilan Eskinazi and Benjamin J Fregly, *Surrogate modeling of deformable joint contact using artificial neural networks.* Medical engineering and physics, 2015. **37**(9): p. 885-891.
102. Margarita Kotti, Lynsey D Duffell, Aldo A Faisal, and Alison H McGregor, *Detecting knee osteoarthritis and its discriminating parameters using random forests.* Medical engineering and physics, 2017. **43**: p. 19-29.
103. Prem N Ramkumar, Sergio M Navarro, Heather S Haeberle, Jaret M Karnuta, Michael A Mont, Joseph P Iannotti, Brendan M Patterson, and Viktor E Krebs, *Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models.* The Journal of arthroplasty, 2019. **34**(4): p. 632-637.
104. Sergio M Navarro, Eric Y Wang, Heather S Haeberle, Michael A Mont, Viktor E Krebs, Brendan M Patterson, and Prem N Ramkumar, *Machine learning and primary total knee arthroplasty: patient forecasting for a patient-specific payment model.* The Journal of arthroplasty, 2018. **33**(12): p. 3617-3623.
105. Prem N Ramkumar, Jaret M Karnuta, Sergio M Navarro, Heather S Haeberle, Richard Iorio, Michael A Mont, Brendan M Patterson, and Viktor E Krebs, *Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: development and validation of a deep learning model.* The Journal of arthroplasty, 2019. **34**(10): p. 2228-2234. e1.
106. Jaret M Karnuta, Sergio M Navarro, Heather S Haeberle, Damien G Billow, Viktor E Krebs, and Prem N Ramkumar, *Bundled care for hip fractures: a machine-learning approach to an untenable patient-specific payment model.* Journal of orthopaedic trauma, 2019. **33**(7): p. 324-330.
107. Heather S Haeberle, J Matt Helm, Sergio M Navarro, Jaret M Karnuta, Jonathan L Schaffer, John J Callaghan, Michael A Mont, Atul F Kamath, Viktor E Krebs, and Prem N Ramkumar, *Artificial intelligence and machine learning in lower extremity arthroplasty: a review.* The Journal of arthroplasty, 2019.
108. Elizabeth W Paxton, Christopher F Ake, Maria CS Inacio, Monti Khatod, Danica Marinac-Dabic, and Art Sedrakyan, *Evaluation of total hip arthroplasty devices using a total joint replacement registry.* pharmacoepidemiology and drug safety, 2012. **21**: p. 53-59.
109. Maria C Inacio, Michelle Lorimer, David C Davidson, Richard N De Steiger, Peter L Lewis, and Stephen E Graves, *What is the risk of revision surgery in hydroxyapatite-coated femoral hip stems? findings from a large national registry.* Clinical Orthopaedics and Related Research, 2018. **476**(12): p. 2353.
110. Danielle A Southern, Peter D Faris, Rollin Brant, P Diane Galbraith, Colleen M Norris, Merril L Knudtson, William A Ghali, and APPROACH Investigators, *Kaplan–Meier methods yielded misleading results in competing risk scenarios.* Journal of clinical epidemiology, 2006. **59**(10): p. 1110-1114.
111. Elizabeth W Paxton, Guy Cafri, Szilard Nemes, Michelle Lorimer, Johan Kärrholm, Henrik Malchau, Stephen E Graves, Robert S Namba, and Ola Rolfson, *An international comparison of THA patients, implants, techniques, and survivorship in Sweden, Australia, and the United States.* Acta orthopaedica, 2019. **90**(2): p. 148-152.

112. R De Steiger, M Lorimer, and SE Graves, *Cross-linked polyethylene for total hip arthroplasty markedly reduces revision surgery at 16 years.* JBJS, 2018. **100**(15): p. 1281-1288.
113. Richard de Steiger, Andrea Peng, Peter Lewis, and Stephen Graves, *What is the long-term survival for primary THA with small-head metal-on-metal bearings?* Clinical Orthopaedics and Related Research, 2018. **476**(6): p. 1231.
114. Stein Atle Lie, Nicole Pratt, Philip Ryan, Lars B Engesæter, Leif I Havelin, Ove Furnes, and Stephen Graves, *Duration of the increase in early postoperative mortality after elective hip and knee replacement.* JBJS, 2010. **92**(1): p. 58-63.
115. PL Lewis, SE Graves, RN de Steiger, DG Campbell, Y Peng, A Hatton, and M Lorimer, *Does Knee Prosthesis Survivorship Improve When Implant Designs Change? Findings from the Australian Orthopaedic Association National Joint Replacement Registry.* Clinical Orthopaedics and Related Research, 2020.
116. Guy Cafri, Stephen E Graves, Art Sedrakyan, Juanjuan Fan, Peter Calhoun, Richard N de Steiger, Alana Cuthbert, Michelle Lorimer, and Elizabeth W Paxton, *Postmarket surveillance of arthroplasty device components using machine learning methods.* Pharmacoepidemiology and drug safety, 2019. **28**(11): p. 1440-1447.
117. Rodney A Gabriel, Beamy S Sharma, Christina N Doan, Xiaoqian Jiang, Ulrich H Schmidt, and Florin Vaida, *A predictive model for determining patients not requiring prolonged hospital length of stay after elective primary total hip arthroplasty.* Anesthesia and Analgesia, 2019. **129**(1): p. 43-50.
118. Manuel Huber, Christoph Kurz, and Reiner Leidl, *Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning.* BMC medical informatics and decision making, 2019. **19**(1): p. 3.
119. Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn, *Bias in random forest variable importance measures: Illustrations, sources and a solution.* BMC bioinformatics, 2007. **8**(1): p. 1-21.
120. Andy Liaw and Matthew Wiener, *Classification and regression by randomForest.* R news, 2002. **2**(3): p. 18-22.
121. Hemant Ishwaran, Udaya B Kogalur, Xi Chen, and Andy J Minn, *Random survival forests for high-dimensional data.* Statistical Analysis and Data Mining: The ASA Data Science Journal, 2011. **4**(1): p. 115-132.
122. Roohollah Milimonfared, Reza H Oskouei, Mark Taylor, and Lucian B Solomon, *An intelligent system for image-based rating of corrosion severity at stem taper of retrieved hip replacement implants.* Medical engineering and physics, 2018. **61**: p. 13-24.
123. Mark Alan Fontana, Stephen Lyman, Gourab K Sarker, Douglas E Padgett, and Catherine H MacLean, *Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty?* Clinical Orthopaedics and Related Research, 2019. **477**(6): p. 1267-1279.
124. Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, and Jorge Cuadros, *Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs.* Jama, 2016. **316**(22): p. 2402-2410.
125. Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang, *Deep learning for health informatics.* IEEE journal of biomedical and health informatics, 2016. **21**(1): p. 4-21.
126. Yann LeCun, Yoshua Bengio, and Geoffrey %J nature Hinton, *Deep learning.* nature 2015. **521**(7553): p. 436.
127. Federico Cabitza, Raffaele Rasoini, and Gian Franco Gensini, *Unintended consequences of machine learning in medicine.* Jama, 2017. **318**(6): p. 517-518.
128. Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. *Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission*. in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015. ACM.
129. Riyi Qiu, Yugang Jia, Fei Wang, Pramod Divakarmurthy, Samuel Vinod, Behlool Sabir, and Mirsad Hadzikadic, *Predictive Modeling of the Total Joint Replacement Surgery Risk: a Deep Learning*

Based Approach with Claims Data. AMIA Summits on Translational Science Proceedings, 2019. **2019**: p. 562.

130. Edward Choi, Andy Schuetz, Walter F Stewart, and Jimeng Sun, *Using recurrent neural network models for early detection of heart failure onset.* Journal of the American Medical Informatics Association, 2016. **24**(2): p. 361-370.
131. Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel, *Learning to diagnose with LSTM recurrent neural networks.* arXiv preprint arXiv:.03677, 2015.
132. Aditya V Karhade, Joseph H Schwab, and Hany S Bedair, *Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty.* The Journal of arthroplasty, 2019. **34**(10): p. 2272-2277. e1.
133. Alex HS Harris, Alfred C Kuo, Thomas Bowe, Shalini Gupta, David Nordin, and Nicholas J Giori, *Prediction models for 30-day mortality and complications after total knee and hip arthroplasties for veteran health administration patients with osteoarthritis.* The Journal of arthroplasty, 2018. **33**(5): p. 1539-1545.
134. Alex HS Harris, Alfred C Kuo, Yingjie Weng, Amber W Trickey, Thomas Bowe, and Nicholas J Giori, *Can Machine Learning Methods Produce Accurate and Easy-to-use Prediction Models of 30-day Complications and Mortality After Knee or Hip Arthroplasty?* Clinical Orthopaedics and Related Research, 2019. **477**(2): p. 452-460.
135. Szilard Nemes, Ola Rolfson, and Göran Garellick, *Development and validation of a shared decision-making instrument for health-related quality of life one year after total hip replacement based on quality registries data.* Journal of evaluation in clinical practice, 2018. **24**(1): p. 13-21.
136. Luke J Saunders, Richard A Russell, and David P Crabb, *The coefficient of determination: what determines a useful R2 statistic?* Investigative ophthalmology & visual science, 2012. **53**(11): p. 6830-6832.
137. MT Sanchez-Santos, C Garriga, A Judge, RN Batra, AJ Price, AD Liddle, MK Javaid, C Cooper, DW Murray, and NK Arden, *Development and validation of a clinical prediction model for patient-reported pain and function after primary total knee replacement surgery.* Scientific reports, 2018. **8**(1): p. 3381.
138. V Mezhov, LL Laslett, H Ahedi, CL Blizzard, RM Aspden, JS Gregory, FR Saunders, S Graves, M Lorimer, and IP Munogoda, *Predictors of total hip replacement in community based older adults: a cohort study.* Osteoarthritis and Cartilage, 2019. **27**: p. S79-S80.
139. MCS Inacio, NL Pratt, EE Roughead, and SE Graves, *Evaluation of three co-morbidity measures to predict mortality in patients undergoing total joint arthroplasty.* Osteoarthritis and Cartilage, 2016. **24**(10): p. 1718-1726.
140. Lee Bayliss and Luke D Jones, *The role of artificial intelligence and machine learning in predicting orthopaedic outcomes.* The Bone and Joint Journal, 2019. **101**(12): p. 1476-1478.
141. D Douglas Miller and Eric W Brown, *Artificial intelligence in medical practice: the question to the answer?* The American journal of medicine, 2018. **131**(2): p. 129-133.
142. Ziad Obermeyer and Ezekiel J Emanuel, *Predicting the future—big data, machine learning, and clinical medicine.* The New England journal of medicine, 2016. **375**(13): p. 1216.
143. Jonathan H Chen and Steven M Asch, *Machine learning and prediction in medicine—beyond the peak of inflated expectations.* The New England journal of medicine, 2017. **376**(26): p. 2507.
144. Theodore J Iwashyna and Vincent Liu, *What's so different about big data?. A primer for clinicians trained to think epidemiologically.* Annals of the American Thoracic Society, 2014. **11**(7): p. 1130-1135.
145. Muin J Khoury and John PA Ioannidis, *Big data meets public health.* Science, 2014. **346**(6213): p. 1054-1055.
146. Andrew L Beam and Isaac S Kohane, *Big data and machine learning in health care.* Jama, 2018. **319**(13): p. 1317-1318.
147. Leonard Shan, B Shan, David Graham, and Akshat Saxena, *Total hip replacement: a systematic review and meta-analysis on mid-term quality of life.* Osteoarthritis and cartilage, 2014. **22**(3): p. 389-406.
148. *The Norwegian Arthroplasty Registry - Annual Report 2020*

149. Keijo Mäkelä and Nils P Hailer, *Different, yet strong together: the Nordic Arthroplasty Register Association (NARA).* Acta Orthopaedica, 2021: p. 1-3.
150. Johan Kärrholm, Cecilia Rogmark, Emma Naucler, Jonatan Nåtman, Johanna Vinblad, Maziar Mohaddes, and Ola Rolfson, *Swedish Hip Arthroplasty Register Annual report 2019*. 2021.
151. M Beck, C Brand, and B Christen, *Swiss national hip & knee joint registry—report 2020.* 2020.
152. Rajan Anand, Stephen E Graves, Richard N de Steiger, David C Davidson, Philip Ryan, Lisa N Miller, and Kara Cashman, *What is the benefit of introducing new hip and knee prostheses?* JBJS, 2011. **93**(Supplement_3): p. 51-54.
153. Jonathan L Conroy, Sarah L Whitehouse, Stephen E Graves, Nicole L Pratt, Philip Ryan, and Ross W Crawford, *Risk factors for revision for early dislocation in total hip arthroplasty.* The Journal of arthroplasty, 2008. **23**(6): p. 867-872.
154. SE Graves, R De Steiger, D Davidson, W Donnelly, S Rainbird, MF Lorimer, KS Cashman, and RJ Vial, *The use of femoral stems with exchangeable necks in primary total hip arthroplasty increases the rate of revision.* The bone & joint journal, 2017. **99**(6): p. 766-773.
155. Jack G Skendzel, J David Blaha, and Andrew G Urquhart, *Total hip arthroplasty modular neck failure.* The Journal of arthroplasty, 2011. **26**(2): p. 338. e1-338. e4.
156. Wayne Hoskins, Xavier Griffin, Alesha Hatton, Richard N. de Steiger, and Roger Bingham, *THA for a Fractured Femoral Neck: Comparing the Revision and Dislocation Rates of Standard-head, Large-head, Dual-mobility, and Constrained Liners.* Clinical Orthopaedics and Related Research®, 2021. **479**(1): p. 72-81.
157. Robert Pivec, Aaron J. Johnson, Simon C. Mears, and Michael A. Mont, *Hip arthroplasty.* The Lancet, 2012. **380**(9855): p. 1768-1777.
158. R Core Team, *R: A language and environment for statistical computing.* 2013.
159. TM Therneau, *A Package for Survival Analysis in R. R package version 3.2-11, 2021*.
160. Alboukadel Kassambara, Marcin Kosinski, Przemyslaw Biecek, and S Fabian, *Package 'survminer'.* Drawing Survival Curves using 'ggplot2'(R package version 0.4.9), 2021.
161. Nils P Hailer, *The innovation trap*. 2016, Taylor & Francis.
162. Alison J Smith, Paul Dieppe, Kelly Vernon, Martyn Porter, and Ashley W Blom, *Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales.* The Lancet, 2012. **379**(9822): p. 1199-1204.
163. Brent M Ardaugh, Stephen E Graves, and Rita F Redberg, *The 510 (k) ancestry of a metal-on-metal hip implant.* New England Journal of Medicine, 2013. **368**(2): p. 97-100.
164. B Ravi, R Croxford, and GA Hawker, *Low surgeon volume is associated with increased complications following total hip arthroplasty, after accounting for experience.* Osteoarthritis and Cartilage, 2013. **21**: p. S135-S136.
165. Richard N de Steiger, Lisa N Miller, David C Davidson, Philip Ryan, and Stephen E Graves, *Joint registry approach for identification of outlier prostheses.* Acta orthopaedica, 2013. **84**(4): p. 348-352.
166. *Australian Orthopaedic Association National Joint Replacement Registry. Hip, knee & shoulder arthroplasty: annual report 2008.*
167. Art Sedrakyan, Bruce Campbell, Stephen Graves, and Jack L Cronenwett, *Surgical registries for advancing quality and device surveillance.* The Lancet, 2016. **388**(10052): p. 1358-1360.
168. Jignesh S Shah and William H Maisel, *Recalls and safety alerts affecting automated external defibrillators.* JAMA, 2006. **296**(6): p. 655-660.
169. Frederic S Resnic, *Postmarketing surveillance of medical devices--filling in the gaps.* The New England journal of medicine, 2012. **366**(10): p. 875.
170. *Swedish Hip Arthroplasty Register. Swedish Hip Arthroplasty Register Annual report. 2019.*
171. Andrew A Guccione, David T Felson, Jennifer J Anderson, John M Anthony, Yuqing Zhang, PW Wilson, Margaret Kelly-Hayes, Philip A Wolf, Bernard E Kreger, and William B Kannel, *The effects of specific medical conditions on the functional limitations of elders in the Framingham Study.* American journal of public health, 1994. **84**(3): p. 351-358.
172. Mitchell W Krucoff, Art Sedrakyan, and Sharon-Lise T Normand, *Bridging unmet medical device ecosystem needs with strategically coordinated registries networks.* Jama, 2015. **314**(16): p. 1691-1692.

173. SL Hardoon, JD Lewsey, PJ Gregg, BC Reeves, and JHP van der Meulen, *Continuous monitoring of the performance of hip prostheses.* The Journal of bone and joint surgery. British volume, 2006. **88**(6): p. 716-720.
174. Hemant Ishwaran, Udaya B Kogalur, and Maintainer Udaya B Kogalur, *Package 'randomForestSRC'*. 2021, Version.
175. Leo Breiman, *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
176. Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer, *Random survival forests.* Annals of Applied Statistics, 2008. **2**(3): p. 841-860.
177. Matthias Schmid, Marvin N Wright, and Andreas Ziegler, *On the use of Harrell's C for clinical risk prediction via random survival forests.* Expert Systems with Applications, 2016. **63**: p. 450-459.
178. Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis, *Conditional variable importance for random forests.* BMC bioinformatics, 2008. **9**(1): p. 1-11.
179. Stefan Dietrich, Anna Floegel, Martina Troll, Tilman Kühn, Wolfgang Rathmann, Anette Peters, Disorn Sookthai, Martin Von Bergen, Rudolf Kaaks, and Jerzy Adamski, *Random Survival Forest in practice: a method for modelling complex metabolomics data in time to event analysis.* International journal of epidemiology, 2016. **45**(5): p. 1406-1420.
180. Hemant Ishwaran, Udaya B Kogalur, Eiran Z Gorodeski, Andy J Minn, and Michael S Lauer, *High-dimensional variable selection for survival data.* Journal of the American Statistical Association, 2010. **105**(489): p. 205-217.
181. Guy Cafri, Peter Calhoun, and Juanjuan Fan, *High dimensional variable selection with clustered data: an application of random multivariate survival forests for detection of outlier medical device components.* Journal of Statistical Computation and Simulation, 2019. **89**(8): p. 1410-1422.
182. Trevor Hastie, Robert Tibshirani, and Martin Wainwright, *Statistical learning with sparsity: the lasso and generalizations*. 2019: Chapman and Hall/CRC.
183. Hui Zou and Trevor Hastie, *Regularization and variable selection via the elastic net.* Journal of the royal statistical society: series B (statistical methodology), 2005. **67**(2): p. 301-320.
184. Yoav Benjamini and Yosef Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal statistical society: series B (Methodological), 1995. **57**(1): p. 289-300.
185. Jerome Friedman, Trevor Hastie, and Rob Tibshirani, *Regularization paths for generalized linear models via coordinate descent.* Journal of statistical software, 2010. **33**(1): p. 1.
186. Terry M Therneau and Patricia M Grambsch, *The cox model*, in *Modeling survival data: extending the Cox model*. 2000, Springer. p. 39-77.
187. Stef Van Buuren and Karin %J Journal of statistical software Groothuis-Oudshoorn, *mice: Multivariate imputation by chained equations in R.* 2011. **45**: p. 1-67.
188. David S Siroky, *Navigating random forests and related advances in algorithmic modeling.* Statistics Surveys, 2009. **3**: p. 147-163.
189. David I Broadhurst and Douglas B Kell, *Statistical strategies for avoiding false discoveries in metabolomics and related experiments.* Metabolomics, 2006. **2**(4): p. 171-196.
190. Andrew J Carr, Otto Robertsson, Stephen Graves, Andrew J Price, Nigel K Arden, Andrew Judge, and David J Beard, *Knee replacement.* The Lancet, 2012. **379**(9823): p. 1331-1340.
191. David Conrad and Douglas Dennis, *Patellofemoral Crepitus after Total Knee Arthroplasty: Etiology and Preventive Measures.* Clinics in orthopedic surgery, 2014. **6**: p. 9-19.
192. CEH Scott, E Davidson, DJ MacDonald, TO White, and JF Keating, *Total knee arthroplasty following tibial plateau fracture: a matched cohort study.* The Bone and Joint Journal, 2015. **97**(4): p. 532-538.
193. Keith R Berend, Adolph V Lombardi Jr, and Joanne B Adams, *Total knee arthroplasty in patients with greater than 20 degrees flexion contracture.* Clinical Orthopaedics and Related Research®, 2006. **452**: p. 83-87.
194. Nicholas G Weiss, Javad Parvizi, Robert T Trousdale, Rex D Bryce, and David G Lewallen, *Total knee arthroplasty in patients with a prior fracture of the tibial plateau.* JBJS, 2003. **85**(2): p. 218-221.
195. Isabelle Guyon and André Elisseeff, *An introduction to variable and feature selection.* Journal of machine learning research, 2003. **3**(Mar): p. 1157-1182.

196. J Paul Leigh, *Assessing the importance of an independent variable in multiple regression: is stepwise unwise?* Journal of clinical epidemiology, 1988. **41**(7): p. 669-677.
197. Michael M Ward, Sinisa Pajevic, Jonathan Dreyfuss, and James D Malley, *Short-term prediction of mortality in patients with systemic lupus erythematosus: Classification of outcomes using random forests.* Arthritis Care & Research: Official Journal of the American College of Rheumatology, 2006. **55**(1): p. 74-80.
198. Kathryn L Lunetta, L Brooke Hayward, Jonathan Segal, and Paul Van Eerdewegh, *Screening large-scale association study data: exploiting interactions using random forests.* BMC genetics, 2004. **5**(1): p. 1-13.
199. Peter Söderman, Henrik Malchau, and Peter Herberts, *Outcome after total hip arthroplasty: Part I. General health evaluation in relation to definition of failure in the Swedish National Total Hip Arthroplasty register.* Acta Orthopaedica Scandinavica, 2000. **71**(4): p. 354-359.
200. Kentaro Ise, Keiichi Kawanabe, Jiro Tamura, Haruhiko Akiyama, Koji Goto, and Takashi Nakamura, *Clinical results of the wear performance of cross-linked polyethylene in total hip arthroplasty: prospective randomized trial.* The Journal of arthroplasty, 2009. **24**(8): p. 1216-1220.
201. Joseph A Coory, Kelvin G Tan, Sarah L Whitehouse, Alesha Hatton, Stephen E Graves, and Ross W Crawford, *The outcome of total knee arthroplasty with and without patellar resurfacing up to 17 years: a report from the Australian Orthopaedic Association National Joint Replacement Registry.* The Journal of arthroplasty, 2020. **35**(1): p. 132-138.
202. Colleen A Weeks, Jacquelyn D Marsh, Steven J MacDonald, Stephen Graves, and Edward M Vasarhelyi, *Patellar resurfacing in total knee arthroplasty: a cost-effectiveness analysis.* The Journal of arthroplasty, 2018. **33**(11): p. 3412-3415.
203. Peter Herberts and Henrik Malchau, *How outcome studies have changed total hip arthroplasty practices in Sweden.* Clinical orthopaedics and related research, 1997(344): p. 44-60.
204. Stephen E Graves, David Davidson, Lisa Ingerson, Philip Ryan, Elizabeth C Griffith, Brian FJ McDermott, Heather J McElroy, and Nicole L Pratt, *The Australian orthopaedic association national joint replacement registry.* Medical Journal of Australia, 2004. **180**: p. S31-S34.
205. G Hallan, SA Lie, O Furnes, LB Engesaeter, SE Vollset, and LI Havelin, *Medium-and long-term performance of 11 516 uncemented primary femoral stems from the Norwegian arthroplasty register.* The Journal of bone and joint surgery. British volume, 2007. **89**(12): p. 1574-1580.
206. Jonas Ranstam and Otto Robertsson, *Statistical analysis of arthroplasty register data.* Acta orthopaedica, 2010. **81**(1): p. 10-14.
207. Peter Herberts and Henrik Malchau, *Long-term registration has improved the quality of hip replacement: a review of the Swedish THR Register comparing 160,000 cases.* Acta Orthopaedica Scandinavica, 2000. **71**(2): p. 111-121.
208. Christoph Kolling, Beat R Simmen, G Labek, and Jörg Goldhahn, *Key factors for a successful National Arthroplasty Register.* The Journal of bone and joint surgery. British volume, 2007. **89**(12): p. 1567-1573.
209. *Norwegian Arthroplasty Registry - Annual Report 2019.*
210. Otto Robertsson and Lars Lidgren, *The short-term results of 3 common UKA implants during different periods in Sweden.* The Journal of arthroplasty, 2008. **23**(6): p. 801-807.
211. Birgitte Espehaug, Ove Furnes, Lars B Engesæter, and Leif I Havelin, *18 years of results with cemented primary hip prostheses in the Norwegian Arthroplasty Register: concerns about some newer implants.* Acta orthopaedica, 2009. **80**(4): p. 402-412.
212. Kenneth J Rothman, *Measures of disease frequency.* Modern epidemiology, 1998.
213. Arthur E Hoerl and Robert W %J Technometrics Kennard, *Ridge regression: Biased estimation for nonorthogonal problems.* 1970. **12**(1): p. 55-67.
214. Robert %J Journal of the Royal Statistical Society: Series B Tibshirani, *Regression shrinkage and selection via the lasso.* 1996. **58**(1): p. 267-288.
215. Harald Binder and Martin %J BMC bioinformatics Schumacher, *Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models.* 2008. **9**(1): p. 1-10.
216. Hege M Bøvelstad, Ståle Nygård, Hege L Størvold, Magne Aldrin, Ørnulf Borgan, Arnoldo Frigessi, and Ole Christian %J Bioinformatics Lingjærde, *Predicting survival from microarray data—a comparative study.* 2007. **23**(16): p. 2080-2087.

217. Kai Kammers, Michel Lang, Jan G Hengstler, Marcus Schmidt, and Jörg %J BMC bioinformatics Rahnenführer, *Survival models with preclustered gene groups as covariates.* 2011. **12**(1): p. 1-12.
218. Manuel López-Ibáñez, Jérémie Dubois-Lacoste, Leslie Pérez Cáceres, Mauro Birattari, and Thomas %J Operations Research Perspectives Stützle, *The irace package: Iterated racing for automatic algorithm configuration.* 2016. **3**: p. 43-58.
219. F Hutter, HH Hoos, and K Leyton-Brown, *Learning and intelligent optimization.* 2011.
220. Patrick J Heagerty and Yingye Zheng, *Survival model predictive accuracy and ROC curves.* Biometrics, 2005. **61**(1): p. 92-105.
221. Huan Liu and Hiroshi Motoda, *Computational methods of feature selection*. 2007: CRC Press.
222. Nathan Mantel, *Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure.* Journal of the American Statistical Association, 1963. **58**(303): p. 690-700.
223. Bernd Bischl, Olaf Mersmann, Heike Trautmann, and Claus Weihs, *Resampling methods for meta-model validation with recommendations for evolutionary computation.* Evolutionary computation, 2012. **20**(2): p. 249-275.
224. Peter Ellison, Geir Hallan, Paul Johan Høl, Nils Roar Gjerdet, and Leif Havelin, *Coordinating retrieval and register studies improves postmarket surveillance.* Clinical Orthopaedics and Related Research®, 2012. **470**(11): p. 2995-3002.
225. Ines Kutzner, Geir Hallan, Paul Johan Høl, Ove Furnes, Øystein Gøthesen, Wender Figved, and Peter Ellison, *Early aseptic loosening of a mobile-bearing total knee replacement: a case-control study with retrieval analyses.* Acta orthopaedica, 2018. **89**(1): p. 77-83.
226. Jerome H Friedman, *Greedy function approximation: a gradient boosting machine.* Annals of statistics, 2001: p. 1189-1232.
227. Guy Cafri and Barbara A Bailey, *Understanding variable effects from black box prediction: Quantifying effects in tree ensembles using partial dependence.* Journal of Data Science, 2016. **14**(1): p. 67-95.
228. Elizabeth A Stuart, Sue M Marcus, Marcela V Horvitz-Lennon, Robert D Gibbons, Sharon-Lise T Normand, and C Hendricks Brown, *Using non-experimental data to estimate treatment effects.* Psychiatric annals, 2009. **39**(7): p. 719-728.

Appendices - Primary Total Joint Prostheses

Primary Total Hip Prostheses

Device *I*

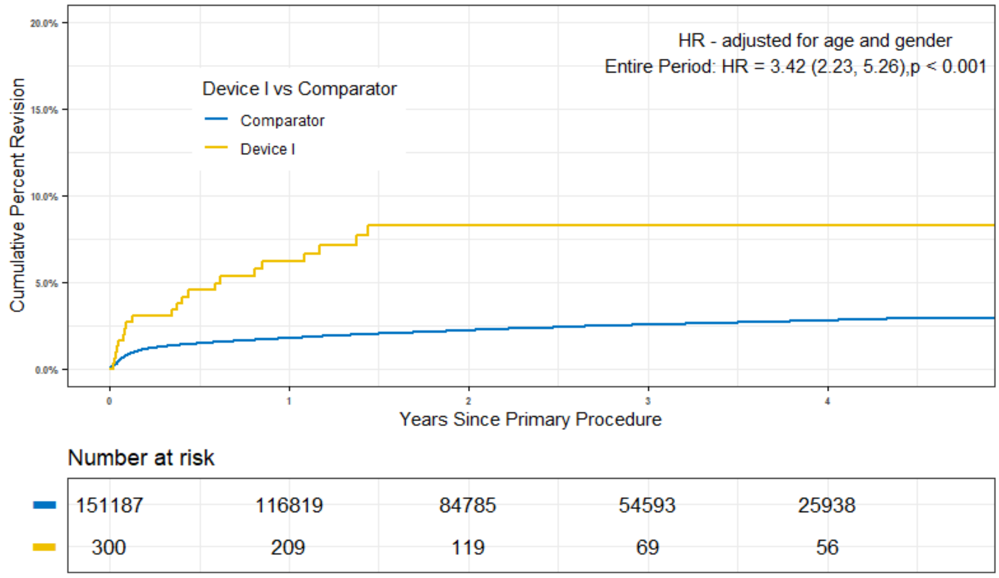


Figure 1. Cumulative percent revision of Device *I*.

Table 1. Revision rates of Device *I* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	21	296	572.6	3.67 (2.27, 5.60)
Hybrid (Femur cemented)	0	4	15	0.00 (0.00, 24.58)
Total	**21**	**300**	**587.6**	**3.57 (3.29, 3.91)**

Table 2. Revision rates of Device *I* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	13	147	296.4	4.39 (2.33, 7.50)
Ceramic/ Non XLPE	7	136	223.7	3.13 (1.26, 6.45)
Metal/Non XLPE	1	17	67.5	1.48 (0.04, 8.25)
Total	**21**	**300**	**587.6**	**3.57 (3.29, 3.91)**

Table 3. Revision rates of Device *I* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	13	158	243.3	5.34 (2.84, 9.14)
Lateral	4	34	101.3	3.95 (1.07, 10.11)
Posterior	4	96	187.7	2.13 (0.58, 5.46)
Total	**21**	**300**	**587.6**	**3.57 (3.29, 3.91)**

Table 4. Reason for revision (follow-up limited to 5 years).

	Device *I*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	5	1.67	23.81
Prosthesis Dislocation	2	0.67	9.52
Fracture	4	1.33	19.05
Loosening	5	1.67	23.81
Leg Length Discrepancy	1	0.33	4.76
Instability	1	0.33	4.76
Pain	-	-	-
Malposition	2	0.67	9.52
Incorrect Sizing	1	0.33	4.76
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	-	-	-
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	-	-	-
N Revision	**21**	**7.0**	**100.0**
N Primary	**300**		

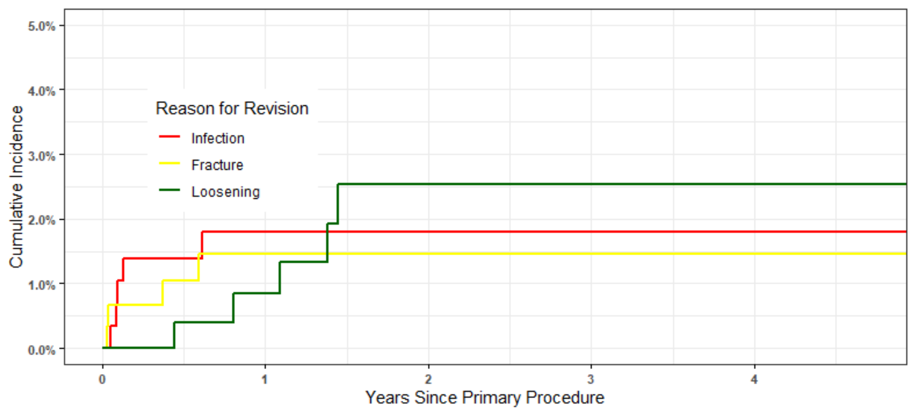


Figure 2. Cumulative incidence revision diagnosis for Device *I*.

Table 5. Type of revision performed for primary total conventional hip replacement.

	Device *I*	
Type of Revision	**Number**	**Percent**
Femoral Component	9	42.85
Acetabular Component	5	23.81
THR (Femoral/Acetabular)	2	9.52
Cement Spacer	2	9.52
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**18**	**85.71**
Head/Insert	2	9.52
Head Only	1	4.76
Minor Components	-	-
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**3**	**14.28**
Total	**21**	**100.0**

Table 6. Revision rates of Device *I* primary total conventional hip replacement by femoral stem.

Femoral stem	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
Exeter V40	0	1	2.8	0.00 (0.00, 132.22)
Furlong	1	18	72.5	1.38 (0.03, 7.68)
Furlong Evolution	18	241	381.7	4.71 (2.79, 7.45)
GTH	0	3	12.2	0.00 (0.00, 30.19)
Linear	2	15	26	7.68 (0.93, 27.73)
Novation	0	21	91.2	0.00 (0.00, 4.04)
Origin	0	1	1.1	0.00 (0.00, 320.77)
Total	**21**	**300**	**587.6**	**3.57 (3.29, 3.91)**

Table 7. Number of revisions of Device *I* primary total conventional hip replacement by year of implant.

Acetabular component	**N Revised**	**N Total**
2015	3	63
2016	0	11
2017	3	53
2018	9	93
2019	6	80
Total	**21**	**300**

Device *II*

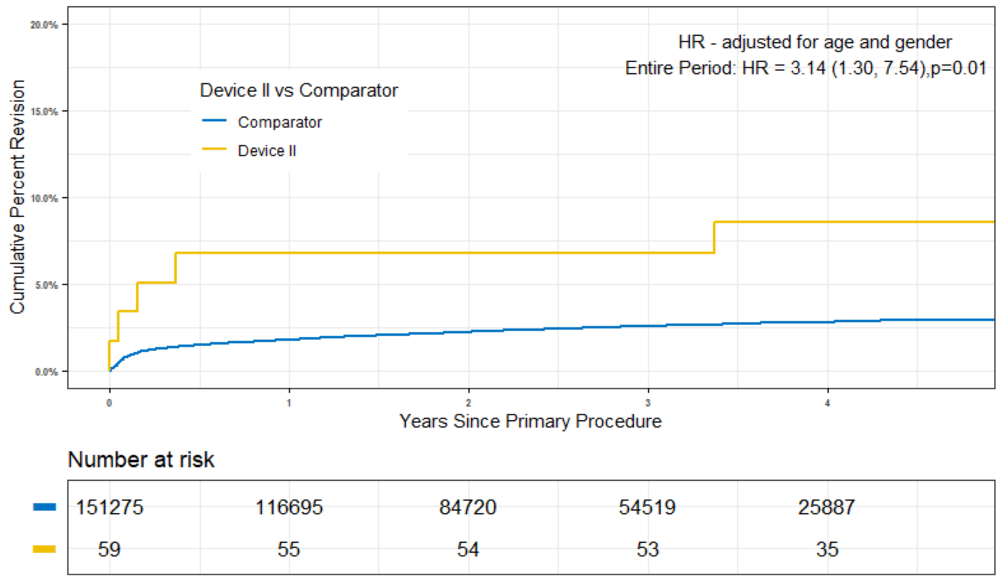


Figure 3. Cumulative percent revision of Device *II*.

Table 8. Revision rates of Device *II* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	5	59	228.8	2.18 (2.03, 2.36)
Total	**5**	**59**	**228.8**	**2.18 (2.03, 2.36)**

Table 9. Revision rates of Device *II* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/XLPE	5	59	228.8	2.18 (2.03, 2.36)
Total	**5**	**59**	**228.8**	**2.18 (2.03, 2.36)**

Table 10. Revision rates of Device *II* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	1	32	128.3	0.78 (0.02, 4.34)
Posterior	1	2	1.9	53.76 (1.36, 299.55)

Table 11. Reason for revision (follow-up limited to 5 years).

	Device *II*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	1	1.7	20.0
Prosthesis Dislocation	-	-	-
Fracture	1	1.7	20.0
Loosening	2	3.4	40.0
Leg Length Discrepancy	1	1.7	20.0
Instability	-	-	-
Pain	-	-	-
Malposition	-	-	-
Incorrect Sizing	-	-	-
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	-	-	-
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	-	-	-
N Revision	**5**	**8.5**	**100.0**
N Primary	**59**		

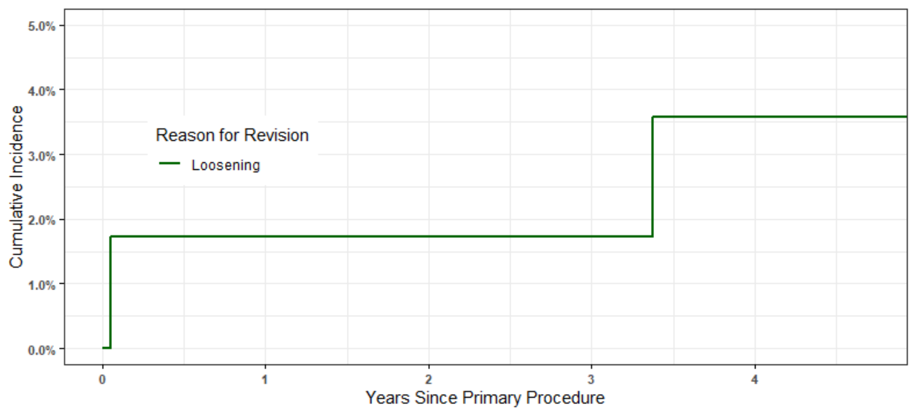


Figure 4. Cumulative incidence revision diagnosis for Device *II*.

Table 12. Type of revision performed for primary total conventional hip replacement.

	Device *II*	
Type of Revision	**Number**	**Percent**
Femoral Component	2	40.0
Acetabular Component	1	20.0
THR (Femoral/Acetabular)	-	-
Cement Spacer	-	-
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**3**	**60.0**
Head/Insert	1	20.0
Head Only	1	20.0
Minor Components	-	-
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**2**	**40.0**
Total	**5**	**100.0**

Table 13. Revision rates of Device *II* primary total conventional hip replacement by femoral stem.

Femoral stem	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Novation stem	5	59	228.8	2.18 (2.03, 2.36)
Total	**5**	**59**	**228.8**	**2.18 (2.03, 2.36)**

Table 14. Number of revisions of Device *II* primary total conventional hip replacement by year of implant.

Acetabular component	N Revised	N Total
2015	3	39
2016	1	18
2017	1	1
2018	0	1
2019	0	0
Total	**5**	**59**

Device *III*

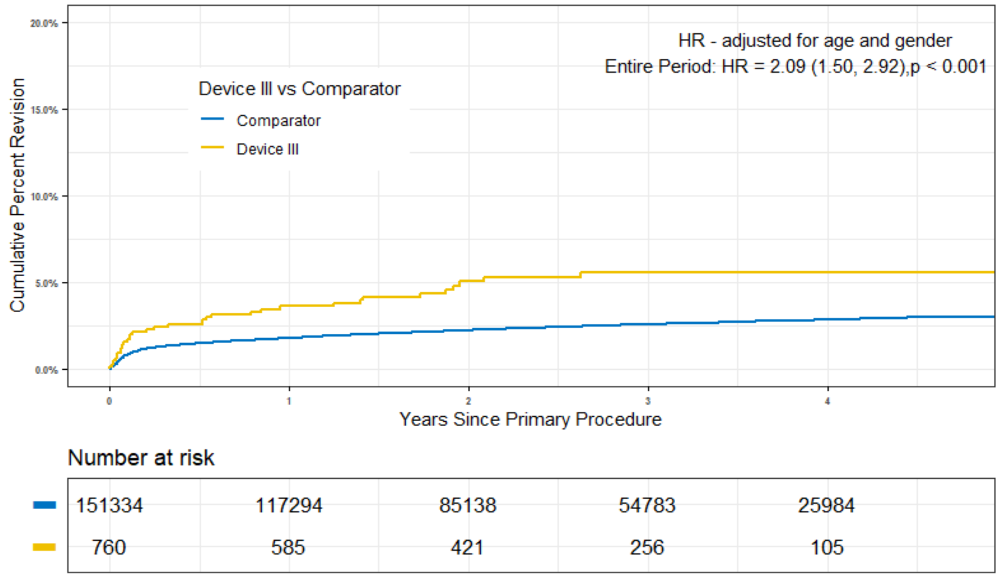


Figure 5. Cumulative percent revision of Device *III*.

Table 15. Revision rates of Device *III* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	34	684	1598.9	2.13 (1.47, 2.97)
Hybrid (Femur cemented)	1	76	136.8	0.73 (0.02, 4.07)
Total	**35**	**760**	**1735.6**	**2.02 (1.93, 2.11)**

Table 16. Revision rates of Device *III* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Non XLPE	31	710	1623.3	1.91 (1.30, 2.71)
Metal/Non XLPE	4	50	112.3	3.56 (0.97, 9.12)
Total	**35**	**760**	**1735.6**	**2.02 (1.93, 2.11)**

Table 17. Revision rates of Device *III* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	26	576	1410.4	1.84 (1.20, 2.70)
Lateral	0	25	47.8	0.00 (0.00, 7.71)
Posterior	8	150	245.3	3.26 (1.41, 6.42)

Table 18. Reason for revision (follow-up limited to 5 years).

	Device *III*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	7	0.92	20.0
Prosthesis Dislocation	2	0.26	5.71
Fracture	13	1.71	37.14
Loosening	7	0.92	20.0
Leg Length Discrepancy	-	-	-
Instability	-	-	-
Pain	1	0.13	2.86
Malposition	1	0.13	2.86
Incorrect Sizing	2	0.26	5.71
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	-	-	-
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	2		
N Revision	**35**	**4.6**	**100**
N Primary	**760**		

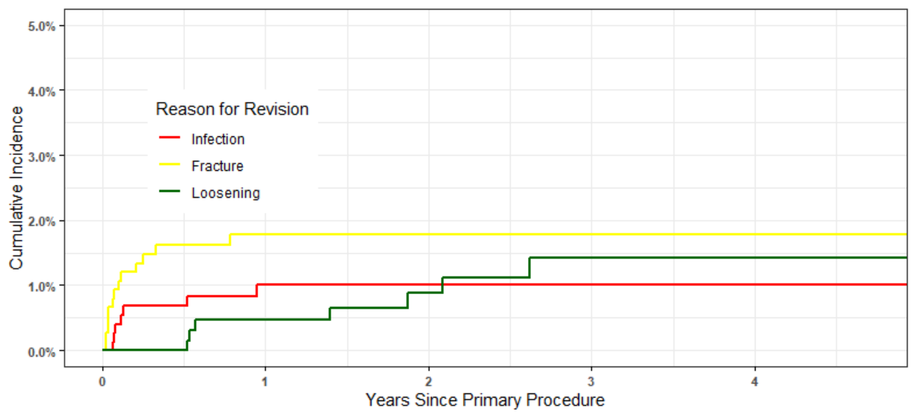


Figure 6. Cumulative incidence revision diagnosis for Device *III*.

Table 19. Type of revision performed for primary total conventional hip replacement.

	Device *III*	
Type of Revision	**Number**	**Percent**
Femoral Component	15	42.86
Acetabular Component	7	20.0
THR (Femoral/Acetabular)	4	11.43
Cement Spacer	1	2.86
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**27**	**77.14**
Head/Insert	7	
Head Only	-	-
Minor Components	1	2.86
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**8**	**22.86**
Total	**35**	**100.0**

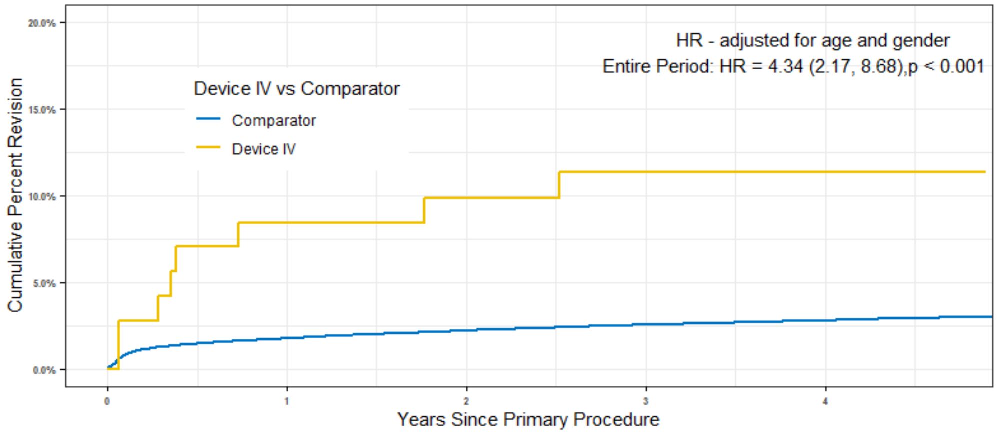
Table 20. Revision rates of Device *III* primary total conventional hip replacement by femoral stem.

Femoral stem	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
AMIStem H	1	35	68.8	1.45 (0.04, 8.09)
CORAIL	0	3	4.4	0.00 (0.00, 83.65)
GHE	2	9	34.7	5.76 (0.70, 20.83)
M/L Taper	0	4	13.7	0.00 (0.00, 26.87)
MasterLoc	0	3	3	0.00 (0.00, 124.20)
MiniMax	5	108	172.7	2.89 (0.94, 6.76)
Modulus	0	1	4.7	0.00 (0.00, 78.32)
Polarstem	0	1	2.4	0.00 (0.00, 154.99)
Quadra-C	0	61	97.9	0.00 (0.00, 3.77)
Quadra-H	26	501	1240.5	2.09 (1.37, 3.07)
Revision Hip	0	2	1.6	0.00 (0.00, 236.47)
S-Rom	0	1	2.8	0.00 (0.00, 129.43)
Taperloc	0	16	49.6	0.00 (0.00, 7.44)
X-Acta	1	15	38.8	2.57 (0.06, 14.34)
Total	**35**	**760**	**1735.6**	**2.02 (1.93, 2.11)**

Table 21. Number of revisions of Device *III* primary total conventional hip replacement by year of implant.

Acetabular component	N Revised	N Total
2015	6	116
2016	8	163
2017	11	173
2018	7	160
2019	3	148
Total	**35**	**760**

Device *IV*



Number at risk

Comparator	151263	116729	84757	54545	25896
Device IV	71	64	63	52	27

Figure 7. Cumulative percent revision of Device *IV*.

Table 22. Revision rates of Device *IV* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	8	71	245.4	3.26 (3.01, 3.56)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 23. Revision rates of Device *IV* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	4	54	191.2	2.09 (0.57, 5.35)
Ceramicised Metal/XLPE	3	12	37.6	7.97 (1.64, 23.30)
Metal/XLPE	1	5	16.5	6.06 (0.15, 33.79)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 24. Revision rates of Device *IV* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Lateral	4	41	140.5	2.85 (0.77, 7.29)
Posterior	3	25	87	3.45 (0.71, 10.07)

Table 25. Reason for revision (follow-up limited to 5 years).

	Device *IV*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	2	2.82	25.0
Prosthesis Dislocation	3	4.22	37.5
Fracture	-	-	-
Loosening	1	1.41	12.5
Leg Length Discrepancy	-	-	-
Instability	-	-	-
Pain	-	-	-
Malposition	-	-	-
Incorrect Sizing	-	-	-
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	1	1.41	12.5
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	1	1.41	12.5
N Revision	**8**	**11.27**	**100.0**
N Primary	**71**		

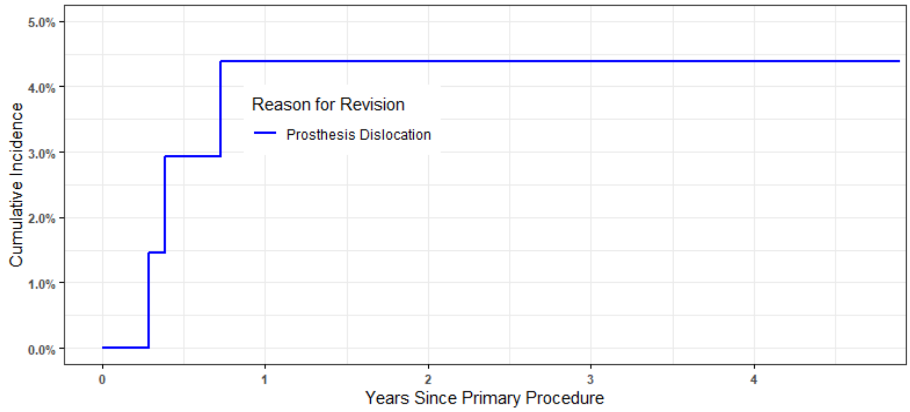


Figure 8. Cumulative incidence revision diagnosis for Device *IV*.

Table 26. Type of revision performed for primary total conventional hip replacement.

	Device *IV*	
Type of Revision	**Number**	**Percent**
Femoral Component	2	25.0
Acetabular Component	-	-
THR (Femoral/Acetabular)	1	12.5
Cement Spacer	-	-
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**3**	**37.5**
Head/Insert	4	50.0
Head Only	1	12.5
Minor Components	-	-
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**5**	**62.5**
Total	**8**	**100.0**

Table 27. Revision rates of Device *IV* primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
R3	8	70	240.6	3.32 (1.43, 6.55)
Trabecular Metal (Shell)	0	1	4.8	0.00 (0.00, 77.17)
Total	**8**	**71**	**245.4**	**3.26 (3.01, 3.56)**

Table 28. Number of revisions of Device *IV* primary total conventional hip replacement by year of implant.

Femoral stem	N Revised	N Total
2015	4	32
2016	4	29
2017	0	10
2018	0	0
2019	0	0
Total	**8**	**71**

Device *V*

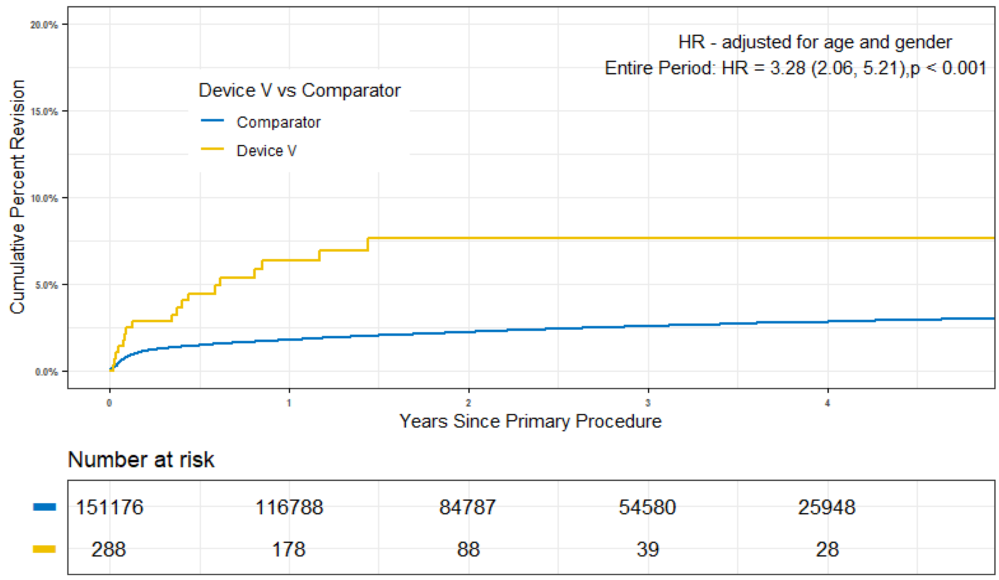


Figure 9. Cumulative percent revision of Device *V*.

Table 29. Revision rates of Device *V* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	18	288	458.7	3.92 (3.59, 4.31)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Table 30. Revision rates of Device *V* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	10	142	235.8	4.24 (2.03, 7.80)
Ceramic/Non XLPE	7	114	129.6	5.40 (2.17, 11.13)
Ceramic/XLPE	0	13	23.9	0.00 (0.00, 15.46)
Ceramic/XLPE + Antioxidant	0	4	12.3	0.00 (0.00, 29.97)
Metal/Non XLPE	1	15	57.2	1.75 (0.04, 9.74)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Table 31. Revision rates of Device *V* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	11	166	230.9	4.76 (2.37, 8.52)
Lateral	4	15	28.3	14.14 (3.85, 36.21)
Posterior	3	100	167.7	1.79 (0.37, 5.23)

Table 32. Reason for revision (follow-up limited to 5 years).

	Device *V*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	5	1.74	27.78
Prosthesis Dislocation	1	0.35	5.55
Fracture	4	1.39	22.22
Loosening	3	1.04	16.67
Leg Length Discrepancy	1	0.35	5.55
Instability	1	0.35	5.55
Pain	-	-	-
Malposition	2	0.69	11.11
Incorrect Sizing	1	0.35	5.55
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	-	-	-
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	-	-	-
N Revision	**18**	**6.25**	**100**
N Primary	**288**		

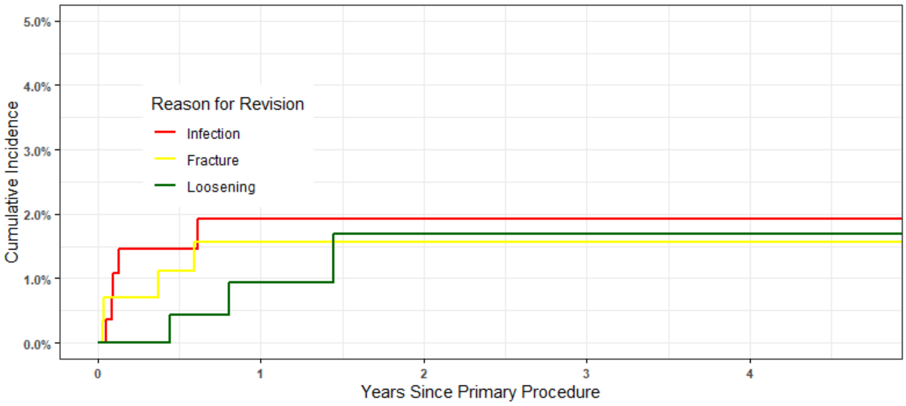


Figure 10. Cumulative incidence revision diagnosis for Device *V*.

Table 33. Type of revision performed for primary total conventional hip replacement.

	Device *V*	
Type of Revision	**Number**	**Percent**
Femoral Component	7	38.89
Acetabular Component	4	22.22
THR (Femoral/Acetabular)	2	11.11
Cement Spacer	2	11.11
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**15**	**83.33**
Head/Insert	2	11.11
Head Only	1	5.55
Minor Components	-	-
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**3**	**16.67**
Total	**18**	**100.0**

Table 34. Revision rates of Device *V* primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Acetabular Shell (Global)	0	4	17.5	0.00 (0.00, 21.11)
Adaptive	0	4	12.3	0.00 (0.00, 29.97)
Delta-TT	0	2	4.3	0.00 (0.00, 85.99)
Furlong	18	241	381.7	4.71 (2.79, 7.45)
Logical G	0	14	24.6	0.00 (0.00, 14.97)
Mpact	0	18	9.1	0.00 (0.00, 40.63)
PINNACLE	0	1	0.2	0.00 (0.00, 1676.76)
Trident/Tritanium (Shell)	0	1	3.4	0.00 (0.00, 109.46)
Versafitcup CC	0	3	5.7	0.00 (0.00, 65.06)
Total	**18**	**288**	**458.7**	**3.92 (3.59, 4.31)**

Table 35. Number of revisions of Device *V* primary total conventional hip replacement by year of implant.

Femoral stem	N Revised	N Total
2015	2	31
2016	0	11
2017	3	52
2018	7	94
2019	6	100
Total	**18**	**288**

Device VI

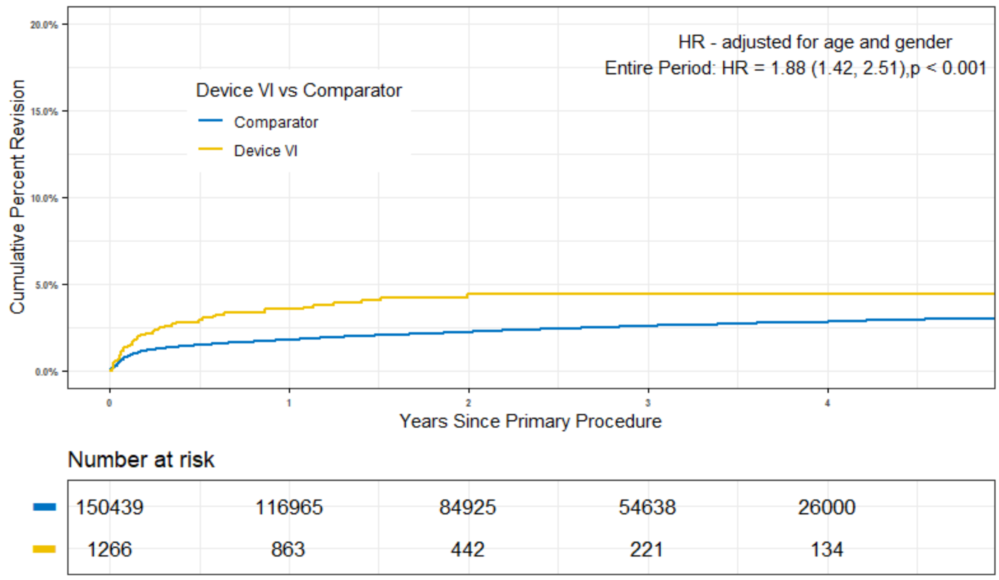


Figure 11. Cumulative percent revision of Device *VI*.

Table 36. Revision rates of Device *VI* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	48	1265	2269.2	2.11 (1.56, 2.80)
Reverse Hybrid (Femur cementless)	0	1	1.8	0.00 (0.00, 203.80)
Total	**48**	**1,266**	**2271**	**2.11 (2.04, 2.2)**

Table 37. Revision rates of Device *VI* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	5	152	403.9	1.24 (0.40, 2.89)
Ceramic/Non XLPE	1	174	400.1	0.25 (0.01, 1.39)
Ceramic/XLPE	38	805	1082.1	3.51 (2.48, 4.82)
Ceramic/XLPE + Antioxidant	1	59	258.6	0.39 (0.01, 2.15)
Metal/Non XLPE	3	39	62.5	4.80 (0.99, 14.04)
Metal/XLPE	0	35	57.2	0.00 (0.00, 6.45)
Metal/XLPE + Antioxidant	0	2	6.7	0.00 (0.00, 55.30)
Total	**48**	**1,266**	**2271**	**2.11 (2.04, 2.2)**

Table 38. Revision rates of Device *VI* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	9	329	551.6	1.63 (0.75, 3.1)
Lateral	1	215	599.3	0.17 (0.00, 0.93)
Posterior	37	698	1023.2	3.62 (2.55, 4.98)

Table 39. Reason for revision (follow-up limited to 5 years).

	Device *VI*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	12	0.95	25.0
Prosthesis Dislocation	10	0.79	20.83
Fracture	16	1.26	33.33
Loosening	3	0.24	6.25
Leg Length Discrepancy	-	-	-
Instability	2	0.16	4.17
Pain	2	0.16	4.17
Malposition	1	0.08	2.08
Incorrect Sizing	-	-	-
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	1	0.08	2.08
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	1	0.08	2.08
N Revision	**48**	**3.79**	**100**
N Primary	**1,266**		

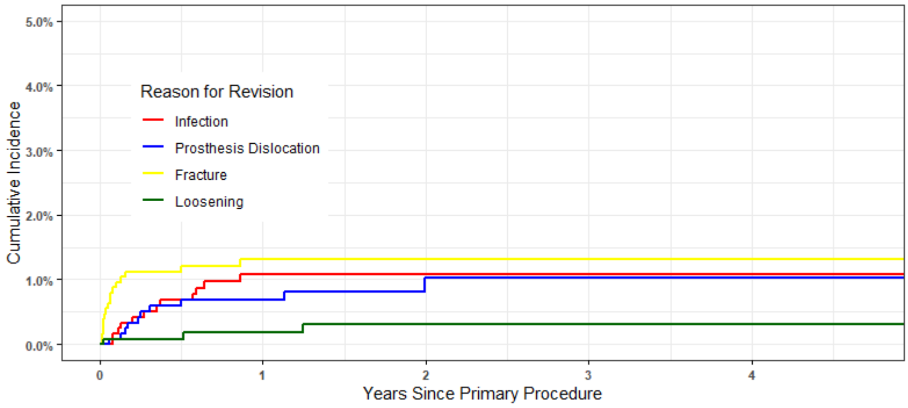


Figure 12. Cumulative incidence of revision diagnosis for Device *VI*.

Table 40. Type of revision performed for primary total conventional hip replacement.

	Device *VI*	
Type of Revision	**Number**	**Percent**
Femoral Component	15	31.25
Acetabular Component	17	0.35
THR (Femoral/Acetabular)	7	0.14
Cement Spacer	2	4.17
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**41**	**85.42**
Head/Insert	6	12.5
Head Only	1	2.08
Minor Components	-	-
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**7**	**14.58**
Total	**48**	**100.0**

Table 41. Revision rates of Device *VI* primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
C2	2	62	165.5	1.2 (0.15, 4.36)
Fin II	0	5	0.6	0 (0.00, 614.81)
Logical G	41	855	1113.1	3.68 (2.64, 5.0)
Marathon	0	1	1.8	0 (0.00, 204.94)
PINNACLE	0	4	18.3	0 (0.00, 20.16)
R3	0	1	2.6	0 (0.00, 144.1)
Saturne	4	213	462.6	0.86 (0.23, 2.21)
Trident/Tritanium (Shell)	0	6	5.1	0 (0.00, 71.91)
Trinity	1	119	501.4	0.2 (0.00, 1.11)
Total	**48**	**1,266**	**2271**	**2.11 (1.56, 2.8)**

Table 42. Number of revisions of Device *VI* primary total conventional hip replacement by year of implant.

Acetabular component	N Revised	N Total
2015	1	144
2016	1	92
2017	15	236
2018	12	419
2019	19	375
Total	**48**	**1,266**

Device *VII*

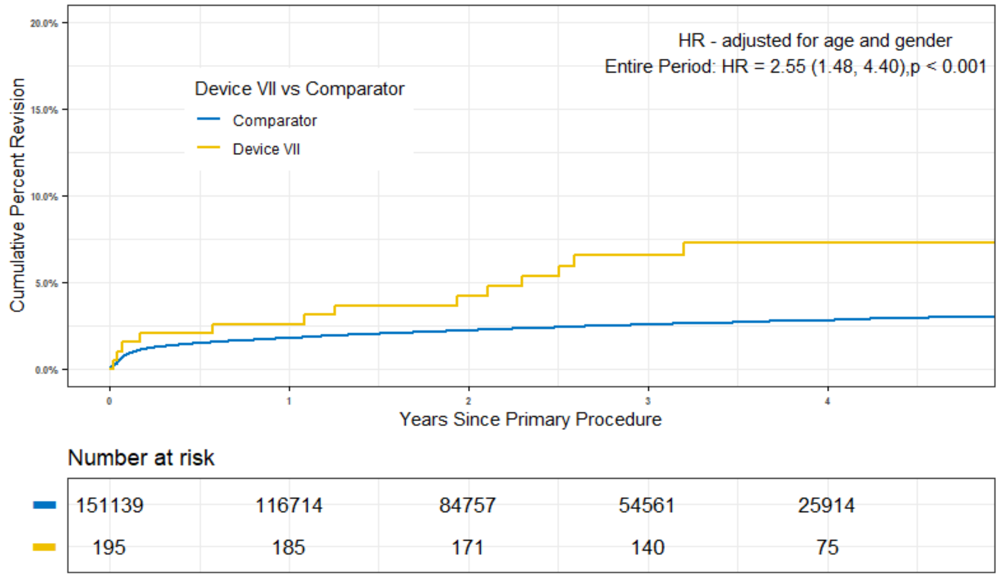


Figure 13. Cumulative percent revision of Device *VII*.

Table 43. Revision rates of Device *VIII* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	13	195	666.5	1.95 (1.86, 2.05)
Total	**13**	**195**	**666.5**	**1.95 (1.86, 2.05)**

Table 44. Revision rates of Device *VII* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	7	93	323.8	2.16 (0.87, 4.45)
Ceramic/XLPE	2	39	135.6	1.47 (0.18, 5.33)
Metal/XLPE	4	63	207.2	1.93 (0.53, 4.94)
Total	**13**	**195**	**666.5**	**1.95 (1.86, 2.05)**

Table 45. Revision rates of Device *VII* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	7	104	365.6	1.91 (0.77, 3.94)
Lateral	2	38	121.2	1.65 (0.20, 5.96)
Posterior	1	23	62.3	1.60 (0.04, 8.94)

Table 46. Reason for revision (follow-up limited to 5 years).

	Device *VII*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	1	0.51	7.69
Prosthesis Dislocation	3	1.54	23.08
Fracture	-	-	-
Loosening	4	2.05	30.77
Leg Length Discrepancy	1	0.51	7.69
Instability	2	1.02	15.38
Pain	1	0.51	7.69
Malposition	1	0.51	7.69
Incorrect Sizing	-	-	-
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	-	-	-
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	-	-	-
N Revision	**13**	**6.67**	**100**
N Primary	**195**		

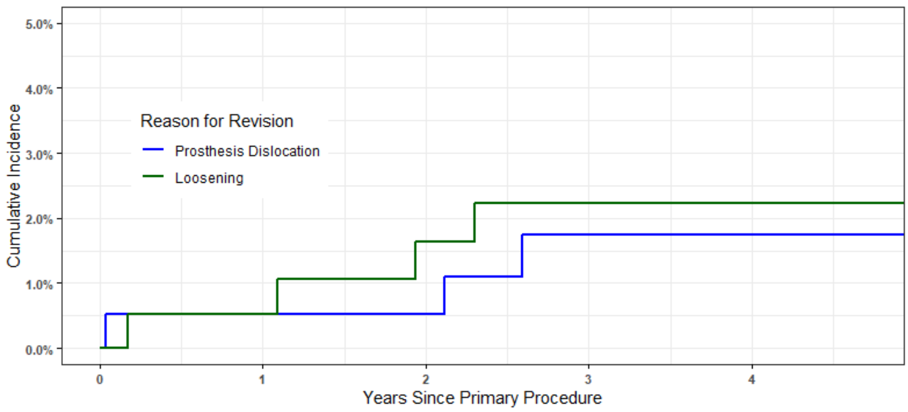


Figure 14. Cumulative incidence revision diagnosis for Device *VII*.

Table 47. Type of revision performed for primary total conventional hip replacement.

	Device *VII*	
Type of Revision	**Number**	**Percent**
Femoral Component	5	38.46
Acetabular Component	3	23.08
THR (Femoral/Acetabular)	3	23.08
Cement Spacer	-	-
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**11**	**84.62**
Head/Insert	2	15.38
Head Only	-	-
Minor Components	-	-
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**2**	**15.38**
Total	**13**	**100.0**

Table 48. Revision rates of Device *VII* primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Acetabular Shell (Global)	7	81	325.2	2.15 (0.86, 4.43)
FMP	3	42	157.8	1.90 (0.39, 5.56)
Furlong	2	15	26.0	7.68 (0.93, 27.73)
Logical G	1	57	157.5	0.63 (0.02, 3.54)
Total	**13**	**195**	**666.5**	**1.95 (1.86, 2.05)**

Table 49. Number of revisions of Device *VII* primary total conventional hip replacement by year of implant.

Femoral stem	N Revised	N Total
2015	5	85
2016	5	68
2017	1	26
2018	2	12
2019	0	4
Total	**13**	**195**

Device VIII

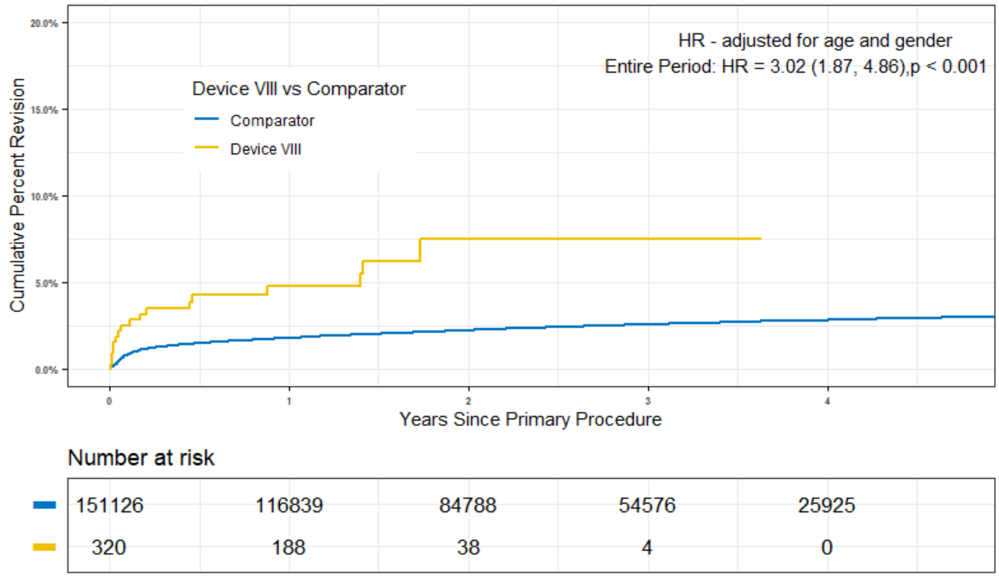


Figure 15. Cumulative percent revision of Device *VIII*.

Table 50. Revision rates of Device *VIII* primary total conventional hip replacement by fixation.

Fixation	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
Cementless	17	320	374.7	4.54 (4.25, 4.87)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Table 51. Revision rates of Device *VIII* primary total conventional hip replacement by bearing surface.

Bearing Surface	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
Ceramic/Ceramic	8	163	161.6	4.89 (2.14, 9.75)
Ceramic/Non XLPE	5	111	179.2	2.79 (0.90, 6.51)
Ceramic/XLPE	3	43	31.8	9.43 (1.94, 27.57)
Metal/Non XLPE	0	1	0.2	0.00 (0.00, 1676.76)
Metal/XLPE	1	2	1.8	55.55 (1.41, 309.53)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Table 52. Revision rates of Device *VIII* primary total conventional hip replacement by approach.

Approach	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Yrs (95% CI)**
Anterior	17	308	355.4	4.78 (2.79, 7.66)
Posterior	0	12	19.2	0.00 (0.00, 19.17)

Table 53. Reason for revision (follow-up limited to 5 years).

	Device *VIII*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	2	0.65	11.76
Prosthesis Dislocation	-	-	-
Fracture	9	2.81	52.94
Loosening	1	0.31	5.88
Leg Length Discrepancy	2	0.62	11.76
Instability	-	-	-
Pain	-	-	-
Malposition	-	-	-
Incorrect Sizing	1	0.31	5.88
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	-	-	-
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	2	0.62	11.76
N Revision	**17**	**5.31**	**100**
N Primary	**320**		

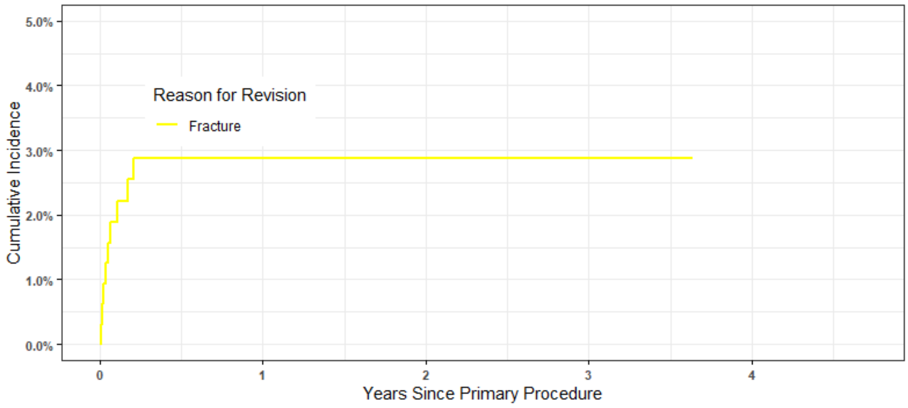


Figure 16. Cumulative incidence revision diagnosis for Device *VIII*.

Table 54. Type of revision performed for primary total conventional hip replacement.

	Device *VIII*	
Type of Revision	**Number**	**Percent**
Femoral Component	9	52.94
Acetabular Component	1	5.88
THR (Femoral/Acetabular)	-	-
Cement Spacer	1	5.88
Removal of Prostheses	-	-
Reinsertion of Components	1	5.88
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**12**	**70.59**
Head/Insert	2	11.76
Head Only	3	17.65
Minor Components	-	-
Insert Only	-	-
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**5**	**29.41**
Total	**17**	**100.0**

Table 55. Revision rates of Device *VIII* primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Agilis	0	1	0.2	0.00 (0.00, 1756.61)
Mpact	8	123	99.5	8.04 (3.47, 15.85)
Trinity	0	1	3.4	0.00 (0.00, 106.92)
Versafitcup CC	4	87	98.9	4.05 (1.10, 10.36)
Versafitcup DM	5	108	172.7	2.89 (0.94, 6.76)
Total	**17**	**320**	**374.7**	**4.54 (4.25, 4.87)**

Table 56. Number of revisions of Device *VIII* primary total conventional hip replacement by year of implant.

Femoral stem	N Revised	N Total
2015	0	0
2016	0	4
2017	3	37
2018	6	155
2019	8	124
Total	**17**	**320**

Device *IX*

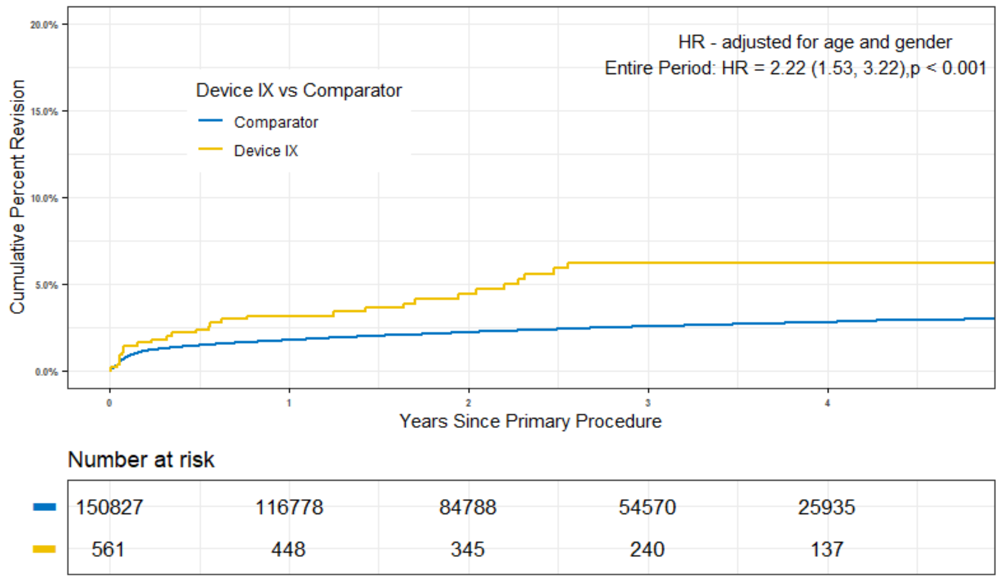


Figure 17. Cumulative percent revision of Device *IX*.

Table 57. Revision rates of Device *IX* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cemented	0	1	0.7	0.00 (0.00, 519.56)
Cementless	28	554	1424	1.97 (1.88, 2.07)
Hybrid (Femur cemented)	0	2	5.6	0.00 (0.00, 65.75)
Reverse Hybrid (Femur cementless)	0	4	8.4	0.00 (0.00, 43.71)
Total	**28**	**561**	**1438.8**	**1.95 (1.29, 2.81)**

Table 58. Revision rates of Device *IX* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	11	181	527.6	1.31 (1.04, 3.73)
Ceramic/Non XLPE	0	4	10.5	0.00 (0.00, 35.2)
Ceramic/XLPE	13	226	605.7	1.22 (1.14, 3.67)
Ceramic/XLPE + Antioxidant	0	50	59.4	0.00 (0.00, 6.21)
Metal/Non XLPE	1	7	16.4	2.88 (0.15, 33.89)
Metal/XLPE	2	85	212.3	0.50 (0.11, 3.4)
Metal/XLPE + Antioxidant	1	8	6.8	2.99 (0.37, 81.34)
Total	**28**	**561**	**1438.8**	**1.95 (1.29, 2.81)**

Table 59. Revision rates of Device *IX* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	1	16	36.6	2.73 (0.07, 15.21)
Lateral	4	108	302.7	1.32 (0.36, 3.38)
Posterior	20	416	1025	4.81 (1.19, 3.01)

Table 60. Reason for revision (follow-up limited to 5 years).

	Device *IX*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	6	1.1	21.4
Prosthesis Dislocation	6	1.1	21.4
Fracture	4	0.7	14.2
Loosening	8	1.4	28.6
Leg Length Discrepancy	1	0.2	3.6
Instability	-	-	-
Pain	-	-	-
Malposition	-	-	-
Incorrect Sizing	-	-	-
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	1	0.2	3.6
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	1	0.2	3.6
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	1	0.2	3.6
N Revision	**28**	**5.1**	**100**
N Primary	**561**		

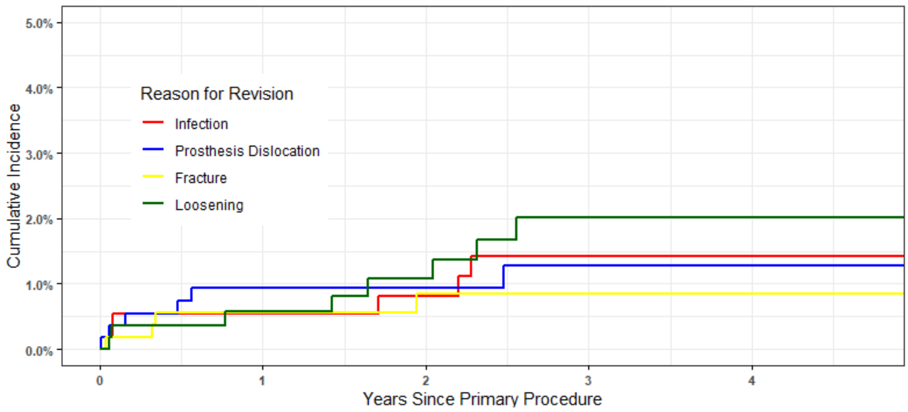


Figure 18. Cumulative incidence of revision diagnosis for Device *IX*.

Table 61. Type of revision performed for primary total conventional hip replacement.

	Device *IX*	
Type of Revision	**Number**	**Percent**
Femoral Component	15	53.6
Acetabular Component	2	7.1
THR (Femoral/Acetabular)	4	14.3
Cement Spacer	2	7.1
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**23**	**82.1**
Head/Insert	4	14.3
Head Only	-	-
Minor Components	-	-
Insert Only	1	3.6
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**5**	**17.9**
Total	**28**	**100.0**

Table 62. Revision rates of Device *IX* primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Acetabular Shell (Global)	0	9	11.6	0.00 (0.00, 31.85)
C2	0	1	3.1	0.00 (0.00, 119)
Continuum	1	3	7.2	13.87 (0.35, 77.28)
Custom Made (Lima)	0	1	2.4	0.00 (0.00, 151.8)
Delta-TT	1	4	1.2	83.33 (2.11, 464.30)
DeltaMotion	1	9	22.7	4.39 (3.23, 6.9)
Dual Mobility Cup	0	2	3.2	0.00 (0.00, 115.64)
G7	1	19	19.1	5.22 (4.02, 7.5)
Mallory-Head	0	1	1.6	0.00 (0.00, 226.31)
Muller	0	2	5.2	0.00 (0.00, 70.26)
Novae	0	4	13.4	0.00 (0.00, 27.43)
PINNACLE	21	414	1163.2	1.8 (1.72, 1.9)
R3	0	4	11.1	0.00 (0.00, 33.23)
Trabecular Metal (Shell)	1	7	21	4.77 (3.38, 8.07)
Trident (Shell)	1	14	35.1	2.85 (2.14, 4.28)
Trident/Tritanium (Shell)	0	4	11	0.00 (0.00, 33.44)
Trinity	1	62	103.6	0.96 (0.81, 1.2)
Versafitcup DM	0	1	2.8	0.00 (0.00, 129.43)
Total	**28**	**561**	**1438.8**	**1.95 (1.29, 2.81)**

Table 63. Number of revisions of Device *IX* primary total conventional hip replacement by year of implant.

Acetabular component	N Revised	N Total
2015	10	153
2016	9	112
2017	3	99
2018	3	102
2019	3	95
Total	**28**	**561**

<u>Device *X*</u>

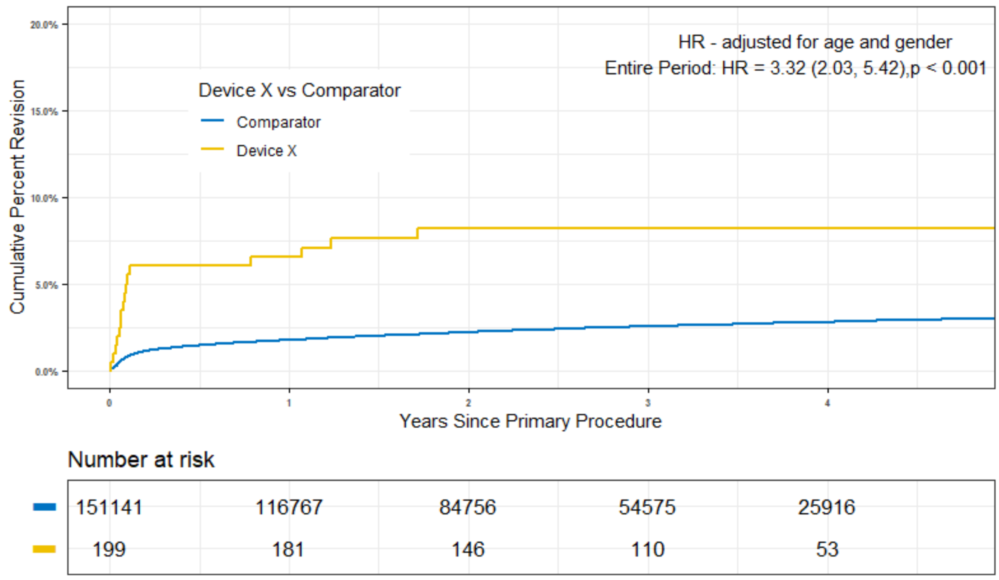


Figure 19. Cumulative percent revision of the Device *X*.

Table 64. Revision rates of Device *X* primary total conventional hip replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cemented	1	4	6	16.52 (8.62, 203.25)
Cementless	15	182	549.3	2.73 (2.56, 2.92)
Hybrid (Femur cemented)	0	13	33.6	0.00 (0.00, 10.96)
Total	**16**	**199**	**589**	**2.72 (1.55, 4.41)**

Table 65. Revision rates of Device *X* primary total conventional hip replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Ceramic/Ceramic	0	1	5	0.00 (0.00, 74.37)
Ceramic/XLPE	0	4	17.8	0.00 (0.00, 20.77)
Ceramic/XLPE + Antioxidant	0	1	5	0.00 (0.00, 74.07)
Metal/Non XLPE	1	3	14.1	7.07 (0.18, 39.40)
Metal/XLPE	15	188	566.3	2.65 (1.48, 4.37)
Metal/XLPE + Antioxidant	0	2	9.9	0.00 (0.00, 37.07)
Total	**16**	**199**	**589**	**2.72 (1.55, 4.41)**

Table 66. Revision rates of Device *X* primary total conventional hip replacement by approach.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Anterior	0	1	4.4	0.00 (0.00, 84.22)
Lateral	2	32	96.8	2.07 (0.25, 7.46)
Posterior	14	150	421.8	3.32 (1.81, 5.57)

Table 67. Reason for revision (follow-up limited to 5 years).

	Device *X*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	2	1.0	12.5
Prosthesis Dislocation	6	3.0	37.5
Fracture	6	3.0	37.5
Loosening	2	1.0	12.5
Leg Length Discrepancy	-	-	-
Instability	-	-	-
Pain	-	-	-
Malposition	-	-	-
Incorrect Sizing	-	-	-
Implant Breakage Acetabular Insert	-	-	-
Implant Breakage Stem	-	-	-
Lysis	-	-	-
Implant Breakage Acetabular	-	-	-
Wear Head	-	-	-
Metal Related Pathology	-	-	-
Wear Acetabular Insert	-	-	-
Implant Breakage Head	-	-	-
Tumour	-	-	-
Heterotopic Bone	-	-	-
Wear Acetabulum	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Progression Of Disease	-	-	-
Other	-	-	-
N Revision	**16**	**8.0**	**100**
N Primary	**199**		

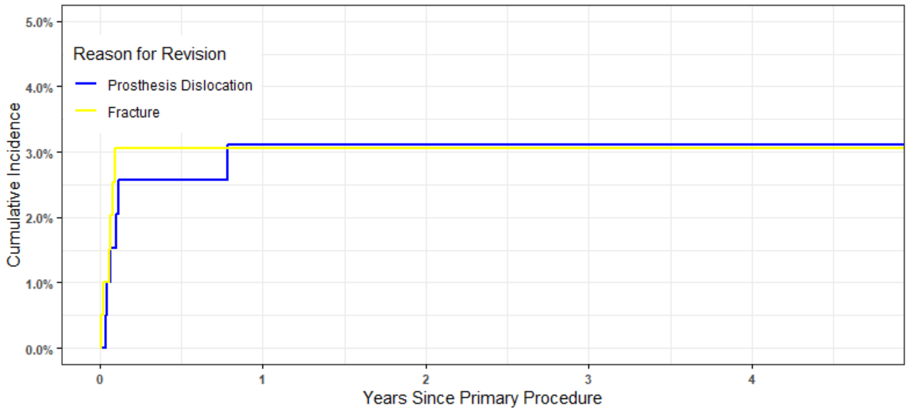


Figure 20. Cumulative incidence revision diagnosis for Device *X*.

Table 68. Type of revision performed for primary total conventional hip replacement.

	Device *X*	
Type of Revision	**Number**	**Percent**
Femoral Component	8	50.0
Acetabular Component	1	6.25
THR (Femoral/Acetabular)	1	6.25
Cement Spacer	2	12.5
Removal of Prostheses	-	-
Reinsertion of Components	-	-
Total Femoral	-	-
Bipolar Head and Femoral	-	-
Saddle	-	-
N Major	**12**	**75.0**
Head/Insert	1	6.25
Head Only	1	6.25
Minor Components	1	6.25
Insert Only	1	6.25
Head/Neck/Insert	-	-
Head/Neck	-	-
Bipolar Only	-	-
Neck Only	-	-
Cement Only	-	-
Neck/Insert	-	-
N Minor	**4**	**25.0**
Total	**16**	**100.0**

Table 69. Revision rates of Device *X* primary total conventional hip replacement by acetabular component.

Acetabular component	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Allofit	0	1	4.9	0.00 (0.00, 74.67)
Continuum	0	2	5.9	0.00 (0.00, 62.1)
Fitmore	0	2	5.8	0.00 (0.00, 63.71)
G7	0	3	2.4	0.00 (0.00, 155.65)
PINNACLE	0	1	3.6	0.00 (0.00, 103.04)
Trabecular Metal (Shell)	0	3	5.3	0.00 (0.00, 69.08)
Trilogy	15	183	555	2.7 (2.54, 2.9)
ZCA	1	4	6	16.52 (8.62, 203.25)
Total	**16**	**199**	**589**	**2.72 (1.55, 4.41)**

Table 70. Number of revisions of Device *X* primary total conventional hip replacement by year of implant.

Acetabular component	N Revised	N Total
2015	6	61
2016	4	60
2017	2	39
2018	3	35
2019	1	4
Total	**16**	**199**

Primary Total Knee Prostheses

Device *I*

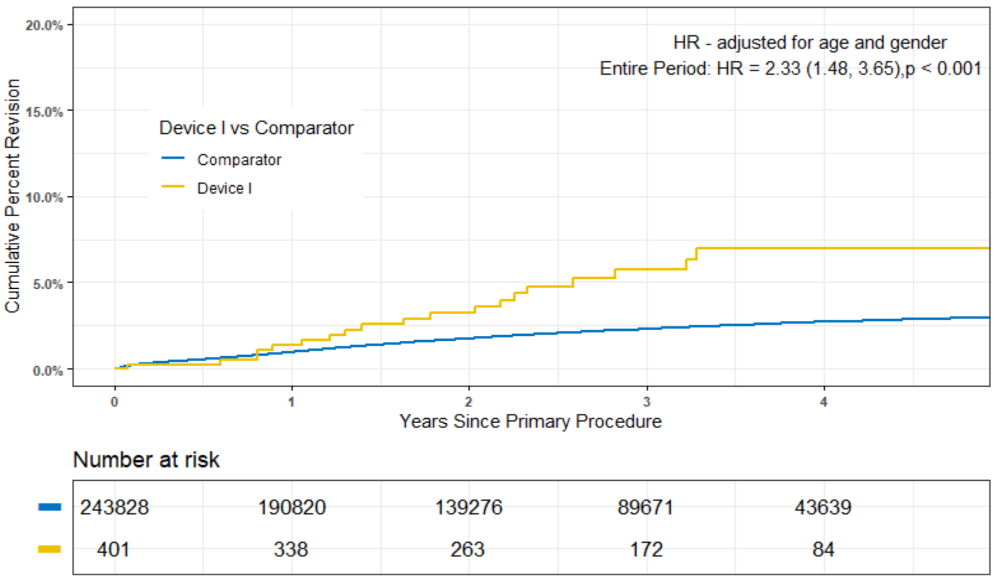


Figure 1. Cumulative percent revision of Device *I*.

Table 1. Revision rates of Device *I* primary total knee replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cementless	19	399	1060.0	1.79 (1.08, 2.80)
Hybrid (Tibial cemented)	0	1	4.7	0.00 (0.00, 77.66)
Hybrid (Tibial cementless)	0	1	4	0.00 (0.00, 92.45)
Total	**19**	**401**	**1068.8**	**1.78 (1.07, 2.78)**

Table 2. Revision rates of Device *I* primary total knee replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Non XLPE	19	401	1068.8	1.78 (1.07, 2.78)
Total	**19**	**401**	**1068.8**	**1.78 (1.07, 2.78)**

Table 3. Revision rates of Device *I* primary total knee replacement by bearing mobility.

Bearing mobility	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Fixed	19	401	1068.8	1.78 (1.07, 2.78)
Total	**19**	**401**	**1068.8**	**1.78 (1.07, 2.78)**

Table 4. Revision rates of Device *I* primary total knee replacement by stability.

Stability	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Medial Pivot Design	19	401	1068.8	1.78 (1.07, 2.78)
Total	**19**	**401**	**1068.8**	**1.78 (1.07, 2.78)**

Table 5. Reason for revision (follow-up limited to 5 years).

	Device *I*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	7	1.7	36.8
Fracture	-	-	-
Loosening	3	0.7	15.8
Instability	3	0.7	15.8
Patella Erosion	-	-	-
Pain	-	-	-
Bearing Dislocation	-	-	-
Malalignment	1	0.2	5.3
Implant Breakage Tibial Insert	-	-	-
Incorrect Sizing	-	-	-
Patellofemoral Pain	1	0.2	5.3
Patella Maltracking	-	-	-
Prosthesis Dislocation	-	-	-
Implant Breakage Femoral	-	-	-
Lysis	-	-	-
Implant Breakage Tibial	-	-	
Heterotopic Bone	-	-	-
Arthrofibrosis	4	1.0	21.1
Wear Tibial Insert	-	-	-
Metal Related Pathology	-	-	-
Implant Breakage Patella	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Wear Patella	-	-	-
Tumour	-	-	-
Wear Tibial	-	-	-
Progression Of Disease	-	-	-
Wear Femoral	-	-	-
Incorrect Side	-	-	-
Post Operative Haematoma	-	-	-
Patella Dislocation	-	-	-
Other	-	-	-
N Revision	**19**	**4.7**	**100.0**
N Primary	**401**		

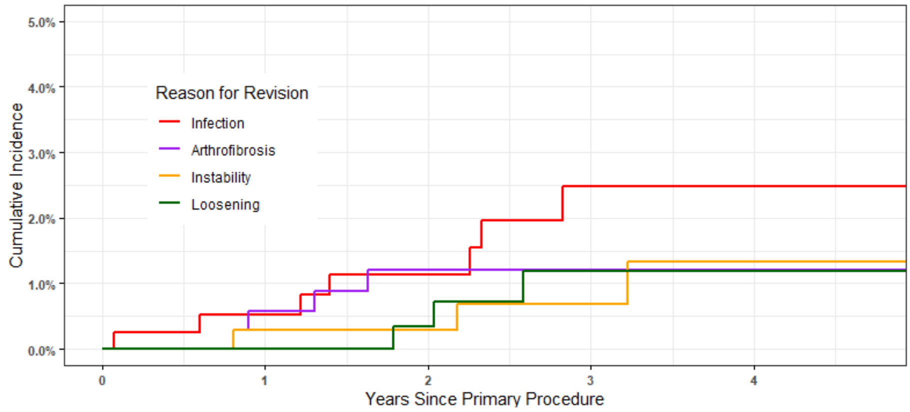


Figure 2. Cumulative incidence revision diagnosis for Device *I*.

Table 6. Type of revision (follow-up limited to 5 years).

	Device *I*	
Type of Revision	**Number**	**Percent**
TKR (Tibial/Femoral)	10	2.5
Tibial Component	-	-
Cement Spacer	-	-
Femoral Component	3	0.7
Removal of Prostheses	-	-
Total Femoral	-	-
Reinsertion of Components	-	-
N Major	**13**	**3.2**
Insert Only	3	0.7
Patella Only	3	0.7
Insert/Patella	-	-
Minor Components	-	-
Cement Only	-	-
N Minor	**6**	**1.4**
Total	**19**	**4.7**

Table 7. Revision rates of Device *I* primary total knee replacement by state.

State	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
NSW	2	134	335.1	0.59 (0.07, 2.16)
SA	17	267	733.7	2.32 (1.35, 3.71)
Total	**19**	**401**	**1068.8**	**1.78 (1.07, 2.78)**

Table 8. Number of revisions of Device *I* primary total knee replacement by year of implant.

Year of implant	N Revised	N Total
2015	6	91
2016	6	97
2017	5	89
2018	1	68
2019	1	56
Total	**19**	**401**

<u>Device *II*</u>

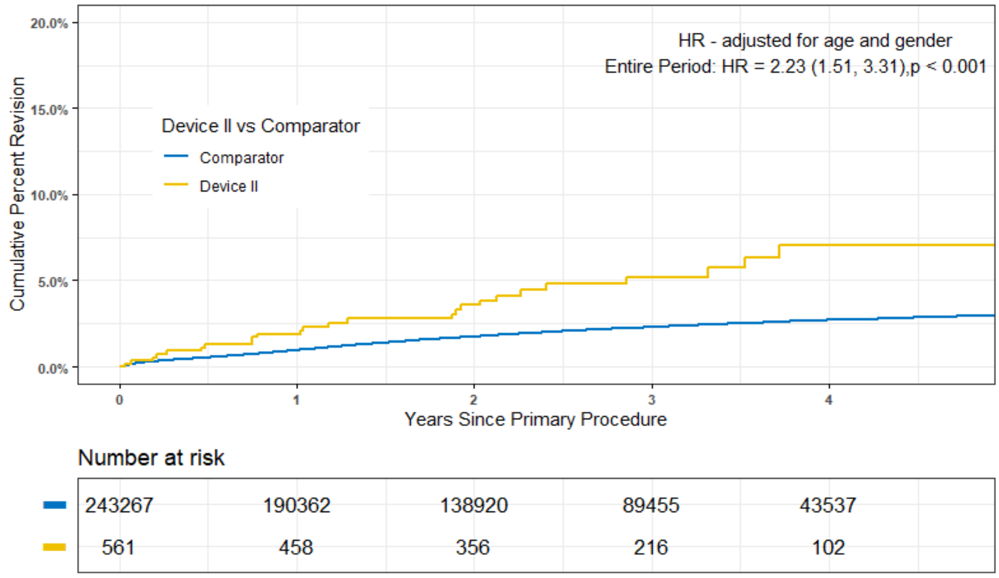


Figure 3. Cumulative percent revision of Device *II*.

Table 9. Revision rates of Device *II* primary total knee replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cemented	14	373	855.3	1.64 (0.89, 2.75)
Cementless	11	128	449.8	2.44 (1.22, 4.37)
Hybrid (Tibial cemented)	0	59	96.5	0.00 (0.00, 3.82)
Hybrid (Tibial cementless)	0	1	4.2	0.00 (0.00, 88.67)
Total	**25**	**561**	**1405.8**	**1.78 (1.15, 2.62)**

Table 10. Revision rates of Device *II* primary total knee replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Non XLPE	25	561	1405.8	1.78 (1.15, 2.62)
Total	**25**	**561**	**1405.8**	**1.78 (1.15, 2.62)**

Table 11. Revision rates of Device *II* primary total knee replacement by bearing mobility.

Bearing mobility	N Revised	N Total	Obs.Years	Revisions/100 Obs.Years (95% CI)
Rotating	25	561	1405.8	1.78 (1.15, 2.62)
Total	**25**	**561**	**1405.8**	**1.78 (1.15, 2.62)**

Table 12. Revision rates of Device *II* primary total knee replacement by stability.

Stability	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Minimally Stabilised	16	303	729.2	2.19 (1.25, 3.56)
Posterior Stabilised	9	258	676.5	1.33 (0.61, 2.52)
Total	**25**	**561**	**1405.8**	**1.78 (1.15, 2.62)**

Table 13. Reason for revision (follow-up limited to 5 years).

	Device *II*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	7	1.2	28.0
Fracture	-	-	-
Loosening	6	1.1	24.0
Instability	-	-	-
Patella Erosion	1	0.2	4.0
Pain	1	0.2	4.0
Bearing Dislocation	-	-	-
Malalignment	1	0.2	4.0
Implant Breakage Tibial Insert	-	-	-
Incorrect Sizing	2	0.4	8.0
Patellofemoral Pain	-	-	-
Patella Maltracking	-	-	-
Prosthesis Dislocation	-	-	-
Implant Breakage Femoral	-	-	-
Lysis	2	0.4	8.0
Implant Breakage Tibial	-	-	-
Heterotopic Bone	-	-	-
Arthrofibrosis	4	0.7	16.0
Wear Tibial Insert	-	-	-
Metal Related Pathology	1	0.2	4.0
Implant Breakage Patella	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Wear Patella	-	-	-
Tumour	-	-	-
Wear Tibial	-	-	-
Progression Of Disease	-	-	-
Wear Femoral	-	-	
Incorrect Side		-	-
Post Operative Haematoma	-	-	-
Patella Dislocation	-	-	-
Other	-	-	-
N Revision	**25**	**4.5**	**100.0**
N Primary	**561**		

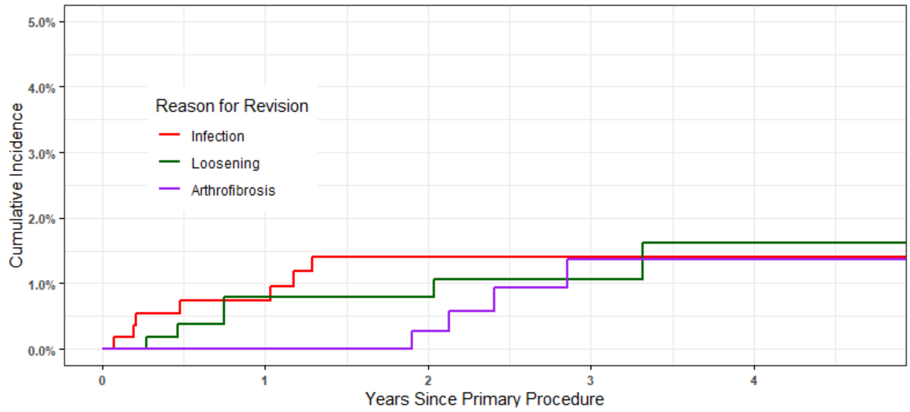


Figure 4. Cumulative incidence revision diagnosis for Device *II*.

Table 14. Type of revision (follow-up limited to 5 years).

	Device *II*	
Type of Revision	**Number**	**Percent**
TKR (Tibial/Femoral)	7	1.2
Tibial Component	1	0.2
Cement Spacer	1	0.2
Femoral Component	4	0.7
Removal of Prostheses	-	-
Total Femoral	-	-
Reinsertion of Components	-	-
N Major	**13**	**2.3**
Insert Only	7	1.2
Patella Only	5	0.9
Insert/Patella	-	-
Minor Components	-	-
Cement Only	-	-
N Minor	**12**	**2.1**
Total	**25**	**4.5**

Table 15. Revision rates of Device *II* primary total knee replacement by state.

State	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
NSW	21	552	1381.6	1.52 (0.94, 2.32)
WA	4	9	24.1	16.57 (4.51, 42.42)
Total	**25**	**561**	**1405.8**	**1.78 (1.24, 2.48)**

Table 16. Number of revisions of Device *II* primary total knee replacement by year of implant.

Year of implant	N Revised	N Total
2015	6	112
2016	14	125
2017	5	140
2018	0	94
2019	0	90
Total	**25**	**561**

Device *III*

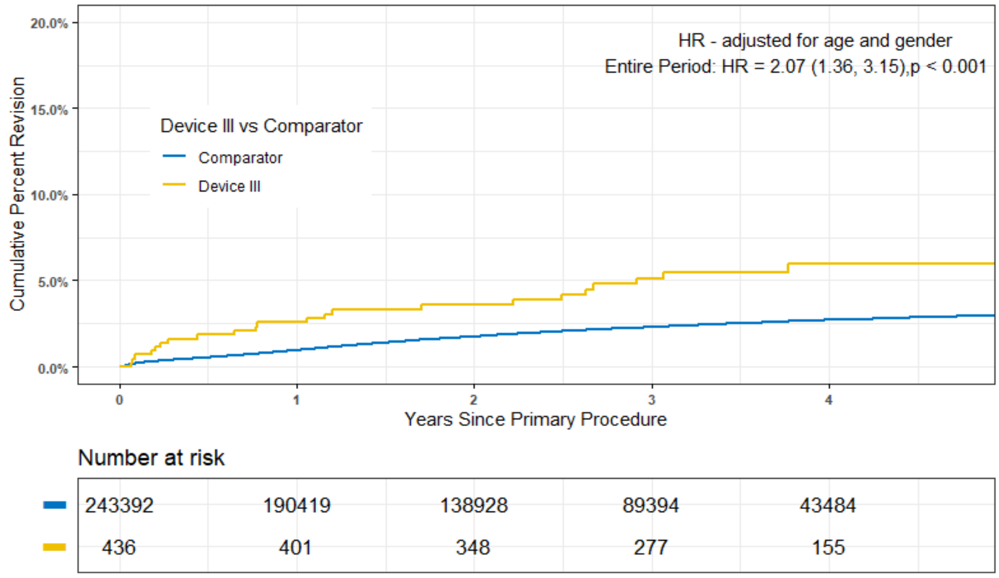


Figure 5. Cumulative percent revision of Device *III*.

Table 17. Revision rates of Device *III* primary total knee replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cemented	22	436	1416.4	1.55 (0.97, 2.35)
Total	**22**	**436**	**1416.4**	**1.55 (0.97, 2.35)**

Table 18. Revision rates of Device *III* primary total knee replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Non XLPE	22	436	1416.4	1.55 (0.97, 2.35)
Total	**22**	**436**	**1416.4**	**1.55 (0.97, 2.35)**

Table 19. Revision rates of Device *III* primary total knee replacement by bearing mobility.

Approach	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Rotating	22	436	1416.4	1.55 (0.97, 2.35)
Total	**22**	**436**	**1416.4**	**1.55 (0.97, 2.35)**

Table 20. Revision rates of Device *III* primary total knee replacement by stability.

Stability	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Posterior Stabilised	22	436	1416.4	1.55 (0.97, 2.35)
Total	**22**	**436**	**1416.4**	**1.55 (0.97, 2.35)**

Table 21. Reason for revision (follow-up limited to 5 years).

	Device *III*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	10	2.3	45.4
Fracture	1	0.2	4.5
Loosening	5	1.1	22.7
Instability	-	-	-
Patella Erosion	-	-	-
Pain	-	-	-
Bearing Dislocation	-	-	-
Malalignment	2	0.5	9.1
Implant Breakage Tibial Insert	-	-	-
Incorrect Sizing	-	-	-
Patellofemoral Pain	2	0.5	9.1
Patella Maltracking	1	0.2	4.5
Prosthesis Dislocation	-	-	-
Implant Breakage Femoral	-	-	-
Lysis	-	-	-
Implant Breakage Tibial	-	-	-
Heterotopic Bone	-	-	-
Arthrofibrosis	1	0.2	4.5
Wear Tibial Insert	-	-	-
Metal Related Pathology	-	-	-
Implant Breakage Patella	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Wear Patella	-	-	-
Tumour	-	-	-
Wear Tibial	-	-	-
Progression Of Disease	-	-	-
Wear Femoral	-	-	-
Incorrect Side	-		-
Post Operative Haematoma	-	-	-
Patella Dislocation	-	-	-
Other	-	-	-
N Revision	**22**	**5.0**	**100.0**
N Primary	**436**		

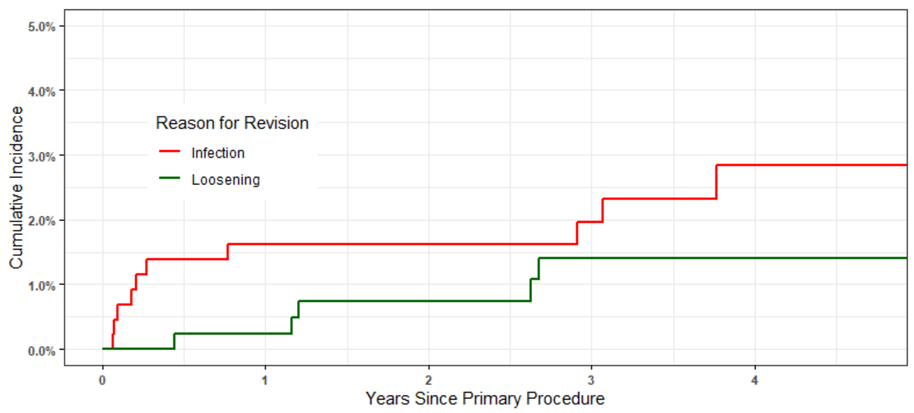


Figure 6. Cumulative incidence revision diagnosis for Device *III*.

Table 22. Type of revision (follow-up limited to 5 years).

	Device *III*	
Type of Revision	**Number**	**Percent**
TKR (Tibial/Femoral)	5	1.1
Tibial Component	3	0.7
Cement Spacer	1	0.2
Femoral Component	1	0.2
Removal of Prostheses	-	-
Total Femoral	-	-
Reinsertion of Components	-	-
N Major	**10**	**2.3**
Insert Only	8	1.8
Patella Only	4	0.9
Insert/Patella	-	-
Minor Components	-	-
Cement Only	-	-
N Minor	**12**	**2.8**
Total	**22**	**5.0**

Table 23. Revision rates of Device *III* primary total knee replacement by state.

State	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
NSW	0	82	242.3	0.00 (0.00, 1.52)
VIC	4	33	85.5	4.68 (1.27, 11.98)
QLD	14	272	968.1	1.45 (0.79, 2.43)
WA	4	41	103.2	3.88 (1.06, 9.92)
SA	0	8	17.4	0.00 (0.00, 21.20)
Total	**22**	**436**	**1416.4**	**1.55 (0.97, 2.35)**

Table 24. Number of revisions of Device *III* primary total knee replacement by year of implant.

Year of implant	**N Revised**	**N Total**
2015	7	165
2016	7	127
2017	5	70
2018	2	50
2019	1	24
Total	**22**	**436**

Device *IV*

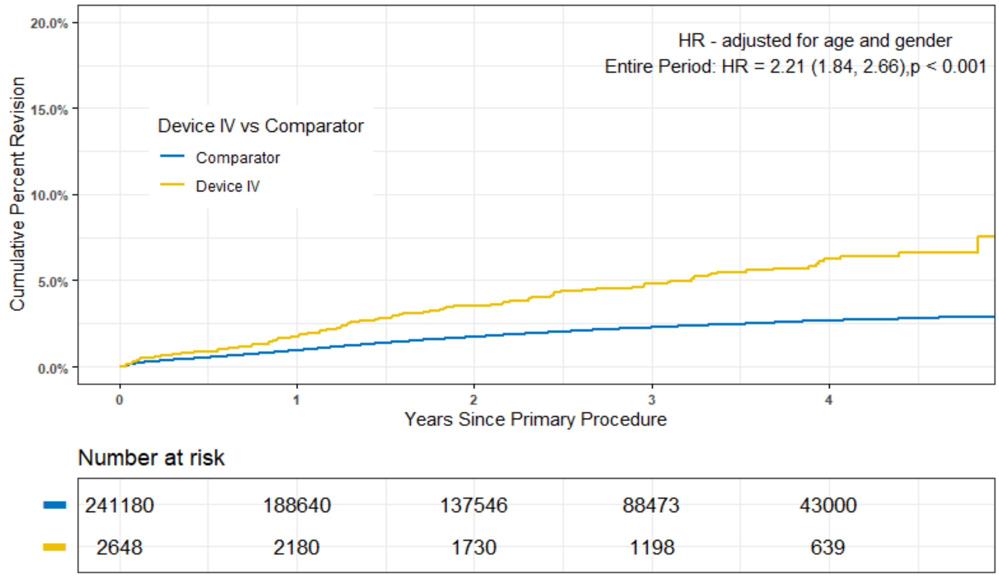


Figure 7. Cumulative percent revision of Device *IV*.

Table 25. Revision rates of Device *IV* primary total knee replacement by fixation.

Fixation	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
Cemented	14	497	1366.9	1.02 (0.56, 1.72)
Cementless	50	932	2390.8	2.09 (1.55, 2.76)
Hybrid (Tibial cemented)	51	1213	3330.4	1.53 (1.14, 2.01)
Hybrid (Tibial cementless)	1	6	21.5	4.65 (0.12, 25.91)
Total	**116**	**2,648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 26. Revision rates of Device *IV* primary total knee replacement by bearing surface.

Bearing Surface	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
Non XLPE	116	2648	7109.6	1.63 (1.35, 1.96)
Total	**116**	**2,648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 27. Revision rates of Device *IV* primary total knee replacement by Bearing mobility.

Bearing mobility	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
Rotating	116	2648	7109.6	1.63 (1.35, 1.96)
Total	**116**	**2,648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 28. Revision rates of Device *IV* primary total knee replacement by stability.

Stability	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
Minimally Stabilised	116	2648	7109.6	1.63 (1.35, 1.96)
Total	**116**	**2,648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 29. Reason for revision (follow-up limited to 5 years).

	Device *IV*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	39	1.5	33.6
Fracture	5	0.2	4.3
Loosening	29	1.1	25.0
Instability	13	0.5	11.2
Patella Erosion	4	0.2	3.4
Pain	12	0.5	10.3
Bearing Dislocation	1	0.0	0.9
Malalignment	4	0.2	3.4
Implant Breakage Tibial Insert	-	-	-
Incorrect Sizing	-	-	-
Patellofemoral Pain	2	0.1	1.7
Patella Maltracking	2	0.1	1.7
Prosthesis Dislocation	-	-	-
Implant Breakage Femoral	-	-	-
Lysis	-	-	-
Implant Breakage Tibial	-	-	-
Heterotopic Bone	-	-	-
Arthrofibrosis	1	0.0	0.9
Wear Tibial Insert	-	-	-
Metal Related Pathology	2	0.1	1.7
Implant Breakage Patella	-	-	-
Synovitis	-	-	
Osteonecrosis	-	-	-
Wear Patella	-	-	-
Tumour	-	-	-
Wear Tibial	-	-	-
Progression Of Disease	-	-	-
Wear Femoral	-	-	-
Incorrect Side	-	-	-
Post Operative Haematoma	-	-	-
Patella Dislocation	-	-	-
Other	2	0.1	1.7
N Revision	**116**	**4.4**	**100.0**
N Primary	**2,648**		

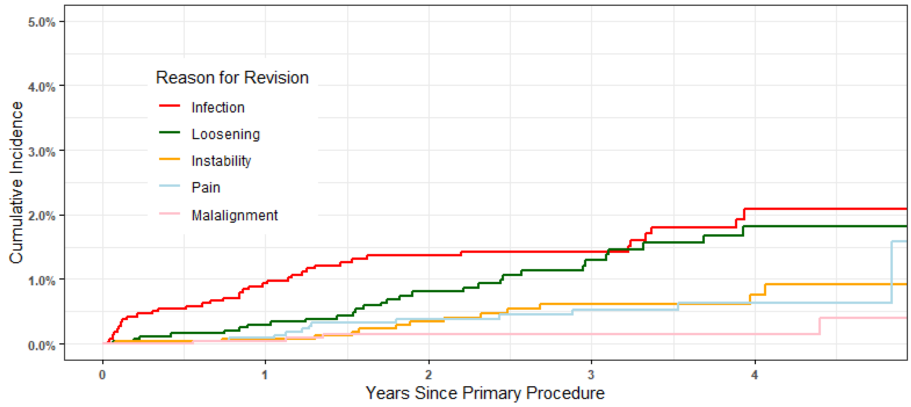


Figure 8. Cumulative incidence revision diagnosis for Device *IV*.

Table 30. Type of revision (follow-up limited to 5 years).

	Device *IV*	
Type of Revision	**Number**	**Percent**
TKR (Tibial/Femoral)	50	1.9
Tibial Component	4	0.2
Cement Spacer	8	0.3
Femoral Component	4	0.2
Removal of Prostheses	-	-
Total Femoral	-	-
Reinsertion of Components	-	-
N Major	**66**	**2.5**
Insert Only	33	1.2
Patella Only	15	0.6
Insert/Patella	2	0.1
Minor Components	-	-
Cement Only	-	-
N Minor	**50**	**1.9**
Total	**116**	**4.4**

Table 31. Revision rates of Device *IV* primary total knee replacement by state.

State	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
NSW	18	578	2189.6	0.82 (0.49, 1.3)
QLD	0	58	183.3	0.00 (0.00, 2.01)
WA	55	1293	2415	2.28 (1.71, 2.96)
SA	43	719	2321.7	1.85 (1.34, 2.49)
Total	**116**	**2,648**	**7109.6**	**1.63 (1.35, 1.96)**

Table 32. Number of revisions of Device *IV* primary total knee replacement by year of implant.

Year of implant	**N Revised**	**N Total**
2015	36	703
2016	34	579
2017	22	527
2018	19	419
2019	5	420
Total	**116**	**2,648**

Device *V*

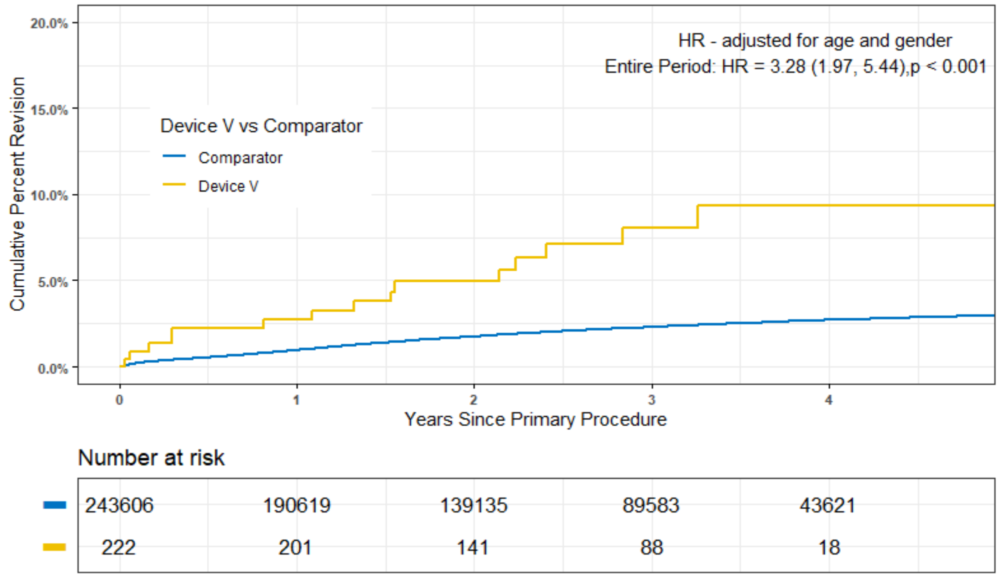


Figure 9. Cumulative percent revision of Device *V*.

Table 33. Revision rates of Device *V* primary total knee replacement by fixation.

Fixation	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Cemented	0	2	4.8	0.00 (0.00, 76.53)
Cementless	2	48	154	1.30 (0.16, 4.69)
Hybrid (Tibial cementless)	13	172	401.2	3.24 (1.72, 5.54)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Table 34. Revision rates of Device *V* primary total knee replacement by bearing surface.

Bearing Surface	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Non XLPE	15	218	543.1	2.76 (1.54, 4.55)
XLPE + Antioxidant	0	4	17	0.00 (0.00, 21.75)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Table 35. Revision rates of Device *V* primary total knee replacement by bearing mobility.

Bearing mobility	N Revised	N Total	Obs.Years	Revisions/100 Obs. Yrs (95% CI)
Fixed	15	222	560.1	2.68 (1.50, 4.42)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Table 36. Revision rates of Device *V* primary total knee replacement by bearing mobility.

Bearing mobility	N Revised	N Total	Obs.Years	Revisions/100 Obs. Years (95% CI)
Posterior Stabilised	15	222	560.1	2.68 (1.50, 4.42)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Table 37. Reason for revision (follow-up limited to 5 years).

	Device *V*		
Revision diagnosis	**Number**	**% Primaries revised**	**% Revisions**
Infection	7	3.2	46.7
Fracture	1	0.5	6.7
Loosening	6	2.7	40.0
Instability	-	-	-
Patella Erosion	-		-
Pain	-		-
Bearing Dislocation	-		-
Malalignment	-	-	-
Implant Breakage Tibial Insert	-	-	-
Incorrect Sizing	-	-	-
Patellofemoral Pain	-	-	-
Patella Maltracking	-	-	-
Prosthesis Dislocation	-	-	--
Implant Breakage Femoral	-	-	-
Lysis	-	-	-
Implant Breakage Tibial	-	-	-
Heterotopic Bone	-	-	-
Arthrofibrosis	1	0.5	6.7
Wear Tibial Insert	-	-	-
Metal Related Pathology	-	-	-
Implant Breakage Patella	-	-	-
Synovitis	-	-	-
Osteonecrosis	-	-	-
Wear Patella	-	-	-
Tumour	-	-	-
Wear Tibial	-	-	-
Progression Of Disease	-	-	-
Wear Femoral	-	-	-
Incorrect Side	-	-	-
Post Operative Haematoma	-	-	-
Patella Dislocation	-	-	-
Other	-	-	-
N Revision	**15**	**6.8**	**100.0**
N Primary	**222**		

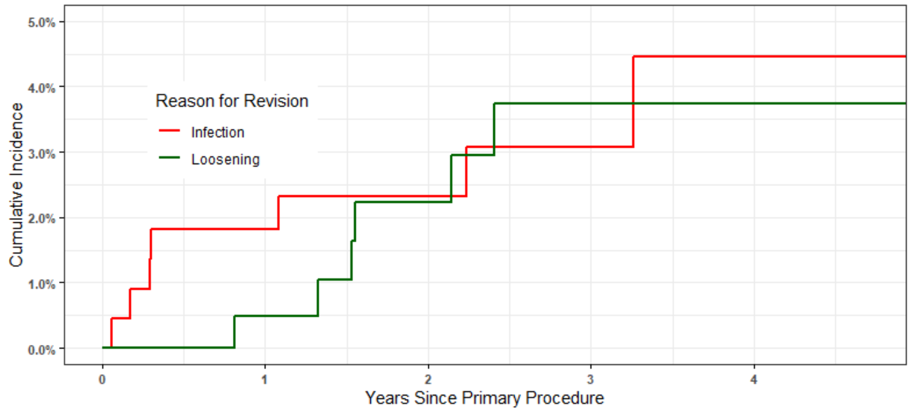


Figure 10. Cumulative incidence revision diagnosis for Device *V*.

Table 38. Type of revision (follow-up limited to 5 years).

	Device *V*	
Type of Revision	**Number**	**Percent**
TKR (Tibial/Femoral)	7	3.2
Tibial Component	-	-
Cement Spacer	3	1.4
Femoral Component	2	0.9
Removal of Prostheses	-	-
Total Femoral	-	-
Reinsertion of Components	-	-
N Major	**12**	**5.4**
Insert Only	3	1.4
Patella Only	-	-
Insert/Patella	-	-
Minor Components	-	-
Cement Only	-	-
N Minor	**3**	**1.4**
Total	**15**	**6.8**

Table 39. Revision rates of Device *V* primary total knee replacement by state.

State	**N Revised**	**N Total**	**Obs.Years**	**Revisions/100 Obs. Years (95% CI)**
NSW	2	47	150.3	1.33 (0.16, 4.81)
VIC	13	175	409.8	3.17 (1.69, 5.42)
Total	**15**	**222**	**560.1**	**2.68 (1.50, 4.42)**

Table 40. Number of revisions of Device *V* primary total knee replacement by year of implant.

Year of implant	**N Revised**	**N Total**
2015	0	18
2016	5	76
2017	10	58
2018	0	56
2019	0	14
Total	**15**	**222**