Identifying optimal femoral neck fractures treatments using a network meta-analysis

Jacob Lavigne

Experimental Surgery

McGill University

September 2018

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Experimental Surgery

Table of Contents

Abstract

French

Les fractures du col du fémur sont des blessures graves caractérisées par un fardeau de maladies important et un taux de mortalité élevé. Le taux de ces blessures est estimé à augmenter et créer un poids considérable pour le secteur de la santé dans les prochaines années. En raison de l'impact de cette blessure sur ceux qui en souffrent ainsi que sur la société, il est nécessaire d'identifier le traitement optimisant la guérison. Pour ce faire, une méta-analyse de réseau a été réalisée selon trois résultats (qualité de vie, mortalité et ostéonécrose avasculaire de la tête du fémur) et les traitements ont été classés en fonction de chaque résultat. De plus, les résultats de l'analyse de la qualité de vie et de la mortalité ont été combinés pour donner un classement sommatif. L'arthroplastie totale de la hanche sans ciment a été classée comme étant le meilleur traitement pour la qualité de vie, tandis que l'hémiarthroplastie unipolaire cimentée était le meilleur en termes de mortalité. Aucun contraste significatif n'a toutefois été trouvé pour cette dernière. La capsulotomie avec réduction ouverte et une greffe osseuse iliaque a été classé la meilleure pour l'ostéonécrose. Les broches étaient significativement supérieures aux vis canulées pour cette complication. Le classement combiné suggère que l'arthroplastie totale avec ou sans ciment ou l'hémiarthroplastie avec ciment devraient être utilisés comme traitement des fractures du col du fémur déplacées chez les personnes âgées. Alors que les deux premiers étaient classés relativement bas (respectivement sixième et troisième) pour la mortalité, ils n'étaient pas significativement pires que les autres traitements. Des recommandations sont présentées pour faire progresser les connaissances sur cette blessure hautement problématique.

English

Femoral neck fractures are serious injuries characterised by a significant burden of illness and an elevated mortality rate. The rate of this injury is forecasted to increase and create a substantial burden on the healthcare sector. Due to the impact this injury has on those who suffer them as well as society, there is a need to identify the treatment that optimizes the healing process. To do so, a network meta-analysis was conducted on three outcomes (quality of life, mortality and avascular necrosis) and treatments were ranked according to each outcome. Furthermore, the results of the quality of life analysis and mortality were combined to give a multi-outcome ranking. Uncemented total hip arthroplasty was ranked best for QoL, whereas cemented unipolar hemiarthroplasty was best for mortality. No significant contrasts were found for mortality however. Capsulotomy, with open reduction an iliac graft was found to be ranked highest for avascular necrosis. Pins were significantly superior to cannulated screws for this outcome. Combined rankings suggest that either total hip arthroplasty with or without cement, or hemiarthroplasty with cement should be used as a treatment for displaced femoral neck fractures in the elderly. While the former two were ranked poorly (sixth and third, respectively) for mortality, they were not significantly worse than other treatments. Recommendations are presented to further advance knowledge for this highly problematic injury.

Acknowledgements

There are numerous individuals that I owe my gratitude to, that helped me throughout my graduate studies and without whom I would not have progressed in my research as I have. Having had to change supervisors early on my studies, I am indebted to my supervisor, Dr. John Sampalis for having taken me up. I am grateful for his guidance, but also a degree of freedom that has allowed my curiosity to express itself. I also owe thanks to all members of my Research Advisory Committee (Dr. Ron Dimentberg, Dr. Robert Turcotte, Dr. John Antoniou, Dr. Olga Huk, Dr. Peter Jarzem, Dr. Maria Petropavlovskaya, and Dr. Renzo Cecere) for finding time to attend meetings and offer feedback on my research. I am also grateful for the help of Christie Schneider (former Director of the Experimental Surgery program) and Sharon Turner (current Director) throughout my studies; they were always available and eager to help answer any question I had regarding the program's requirements, as well as for being morally supportive when I needed it. I would also like to thank Gerta Ruecker and Guido Schwarzer, who developed the R package netmeta that was used in this study, as they were quick to answer several questions; David Wilson, author of the book "Practical Meta-Analysis", for answering my questions expeditiously; Nazi Torabi, the McGill librarian that reviewed my database search strategy; and Jose Luis Ramirez, who served as the second bias assessor. On a more personal note, I wish to also thank my physiotherapists, Laura Abatiello and Alex Lafrance, and their support staff; to the staff at Café Pista, Thomson House and Le Zotique for the great coffee and for being so hospitable and patient with me; and to my friends Victor Frankel, Victoria Gosselin, Vincent Pager-Mercier, Julien Ouellet, Hocine Slimani, Sebastien Davies and Alex Grey for their support and advice.

Lastly, I would like to thank my grandparents; Kathleen Markus and the late Dr. Arthur Markus my parents; Michel Lavigne and Dr. Pamela Markus, and my brother Zacharie Lavigne, for their unconditional love and endless support. I love you big like the sky!

Contribution of authors

A total of 3 individuals we involved in advancing this project. JL, the doctoral student submitting this thesis was responsible for data acquisition (database searching, paper review and selection, data extraction and cleaning), analysis, figure and table generation, and writing, editing and formatting of this thesis. Jose Luis Ramirez conducted the second bias assessment. Dr. John Sampalis was available for advising on various parts of this study.

Introduction

Occurring at the neck, intertrochanteric and subtrochanteric regions of the proximal femur, hip fractures are characterised by substantial morbidity and mortality. These serious injuries are a growing problem for the healthcare system and their treatment must be improved to minimize the personal and social suffering associated with them. This study will identify hip fracture treatments that optimize the quality of life and minimize mortality and occurrence of avascular necrosis of the femoral head in patients suffering femoral neck fractures (FNF), a specific type of hip fracture.

To comprehend the analyses and interpretations covered in this thesis, various principles as well as the framework (research question, hypotheses, etc.) for this study will be discussed first. Topics covered in the background review below include the patient profile, trends in rates and socioeconomics, anatomy of the proximal femur, etiology and fracture patterns, diagnosis and treatments, and burden of illness. This is followed by an outline of this study's rationale, research questions, hypotheses, and objectives.

Background (literature review)

Patient profile

Hip fractures are a significant problem facing our healthcare system and describing the characteristics that contribute to the risk of fractures (risk factors) is an important initial step in addressing this pressing issue. As described elsewhere, these risk factors can be broken down into three categories; those that lower bone strength, increase the risk of falling, and various clinical characteristics that affect the former two [1]. There are numerous hip fracture risk factors, but for simplicity sake, only principle ones are discussed here as the purpose of discussing risk factors is simply to describe the generic profile of hip fracture patients. A more thorough review can be found elsewhere [1, 2].

Bone strength describes the combined effect of bone mineral density (BMD), bone micro- and macro-architecture, bone turnover rate, microdamage, and mineralization. These factors together are known as the physical resilience of bone to mechanical stress and can be affected by various pathologies; the most common is osteoporosis, which can be either primary, secondary or idiopathic [1]. Conditions such as rickets, osteomalacia, renal osteodystrophy, Paget's disease, osteogenesis imperfecta, bone cancers, and other skeletal disorders are also associated with decreased bone strength [3]. An increase in the risk of falling can lead to an increase in fracture risk. Risk factors associated with the risk of falling include impairments in functional mobility, visual or proprioceptive impairments, neurologic conditions, muscular weakness and postural sway [1]. Clinical risk factors include being a Caucasian woman, elderly, certain medication, prior fragility fracture, cigarette smoking, alcohol use, weight, height, low body mass index, inactivity, among various other less-informative factors. Many of these risk factors act on hip facture risk by either reducing bone strength or increasing the risk of falling [1].

Rates and socioeconomic burden

In Quebec, there were approximately 7,500 hip fractures per year between 2000 and 2005 and fracture incidence for women and men in Québec in 2010 was 90.6/100,000 and 56.8/100,000, respectively [4] [5]. The lifetime risk of fracture for a 50-year-old Canadian in

2008 was 12.1% and 4.6% for women and men, respectively [6]. As mentioned above, hip fracture rates are variable among age groups, with the elderly being the most affected. Hence, crude rates can be highly influenced by changing age-distributions of a population. Census data covering the past thirty years show that the population of Québec is aging and will continue to age in coming years. This has led to many experts to predict increasing rates. With direct first year costs of approximately $35,000 (CAD) per patient, hip fractures impose a $650M federal burden, annually, and this sum is predicted to nearly double by 2041 [7]. In Quebec specifically, hip fracture cost an estimated $176M in 2016 [8]. To reduce the costs of treating these injuries, treatments must be optimized. As there are various options available, identifying the optimal treatment is crucial.

Anatomy of the hip

To understand hip fracture patterns, how complications arise, and how treatments can be optimized to better outcomes, a brief review of the anatomy of the hip is presented below. This will be limited to the anatomy necessary to understand the development of complications, principally to the blood supply to the femoral head.

When referring to hip fractures, the hip consists of the proximal femur, from the femoral head to 5cm inferior to the lesser trochanter. It can be separated into three main regions: the femoral head, femoral neck and trochanteric regions. The femoral head and most of the neck reside in the hip joint capsule, terminated laterally by a fold known as the capsular reflecta. The neck is covered with a synovium that lacks a cambium layer and therefore cannot form a periosteal callus after fracture, which is important for fracture healing [9]. Instead, the proximal hip heals by intra-medullary endosteal callus (also known as primary union or creeping substitution). This synovium is rather thin but enlarges at the posterocaudal and cranial portions of the neck. It contains ligaments known as the Weitbrecht retinacula that run along the femoral neck. Here lie essential blood vessels [9].

Entering the capsule at Claffey's point, the posterocranial arteries supply the lateral epiphyseal system with blood, whereas the posterocaudal arteries feed the caudal metaphyseal network (see figure 1.1) [10]. The capital (also known as round) ligament, which joins the acetabulum to the femoral head also contain arterial branches that feed the femoral head [9, 10]. These vessels are fed and drained by the lateral and medial circumflex arteries and veins, respectively [9]. The medial circumflex artery and vein are most important for femoral head vitality; it branches from the deep femoral artery (more common) or the common femoral artery (less common) and feed the posterocranial arteries, and subsequently the lateral epiphyseal network [9, 10]. While the lateral circumflex vessels lie on the anterior portion of the femoral neck, the medial vessels run posterior to the neck [9].

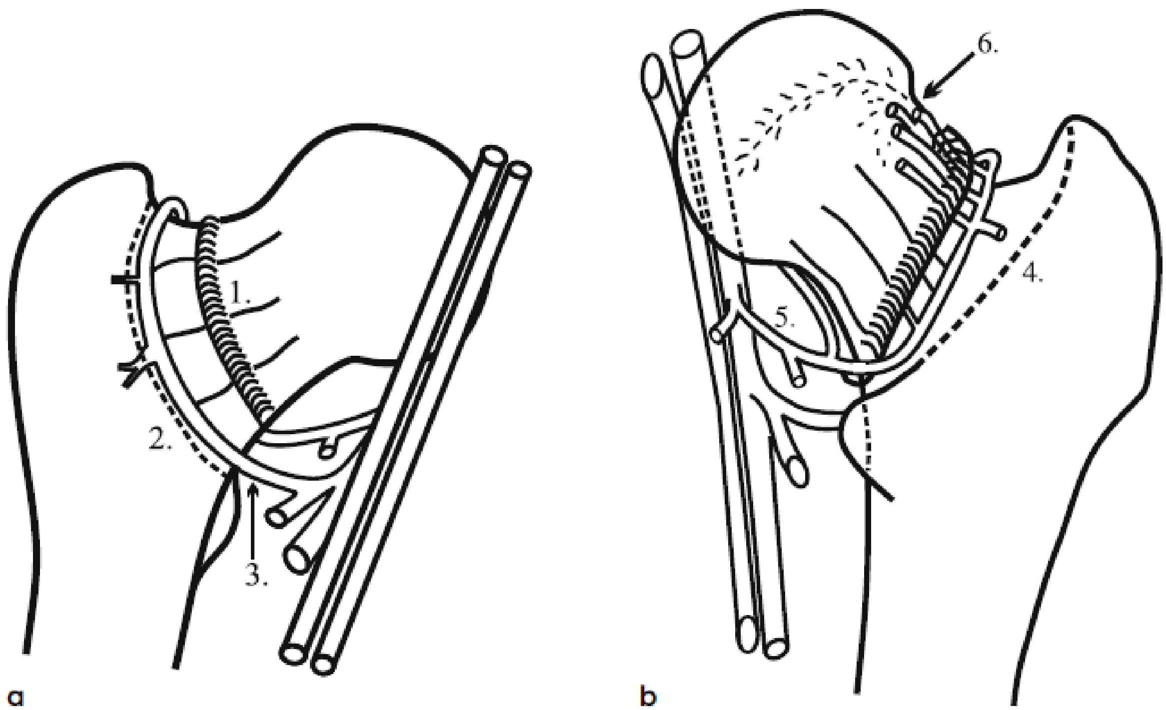


Figure 1.1: Schema of the vasculature of the proximal femur viewed anteriorly (a) and posteriorly (b). The fold of the joint capsule (1), intertrochanteric line (2), anterior circumflex artery (3), intertrochanteric crest (4), medial circumflex artery (5), and Claffey's point (6). Reproduced with permission from Springer (licence number: 4450940728208) [9].

These vessels are of major importance in FNF as they compose the principle blood supply to the femoral head. Experts believe that the disruption of blood flow through them is most responsible for osteonecrosis of the femoral head following FNF [9]. While tears of these ligaments cannot be reversed, torsion, kinking or bruising can be fixed. As blood vessel feeding the femoral head enter the femoral neck bone at Claffey's point, fractures that occur medially to Claffey's point will most have ruptured blood vessels and subsequently experience an increased risk of AVN.

The femoral head is nearly spherical and made up of dense subchondral bone. Two thirds of the femoral head are covered with hyaline cartilage. The fovea capitis does not have cartilage and is where the capital femoral ligament joins it to the acetabular fossa. The Adam's Arch is a thick portion of cortical bone that extends from the lesser trochanter to the femoral head. It decreases in thickness from the trochanter to the head and ends cervically with compression trabeculae.

The calcar femorale is a thick intraosseous bone plate that originates from the caudal femoral neck and together with the Adam's Arch forms a U-shaped gutter when viewed laterally [9]. Some refer to the calcar femorale and Adam's Arch together as the calcar. The gutter plays an important structure in the stability of internal fixation implants for FNF. Together with the subchondral bone, internal fixation is optimized by what is called three-point buttressing (see figure 1.2); mechanical support of the femoral head, calcar and lateral cortex offer optimal internal fixation stability [11].

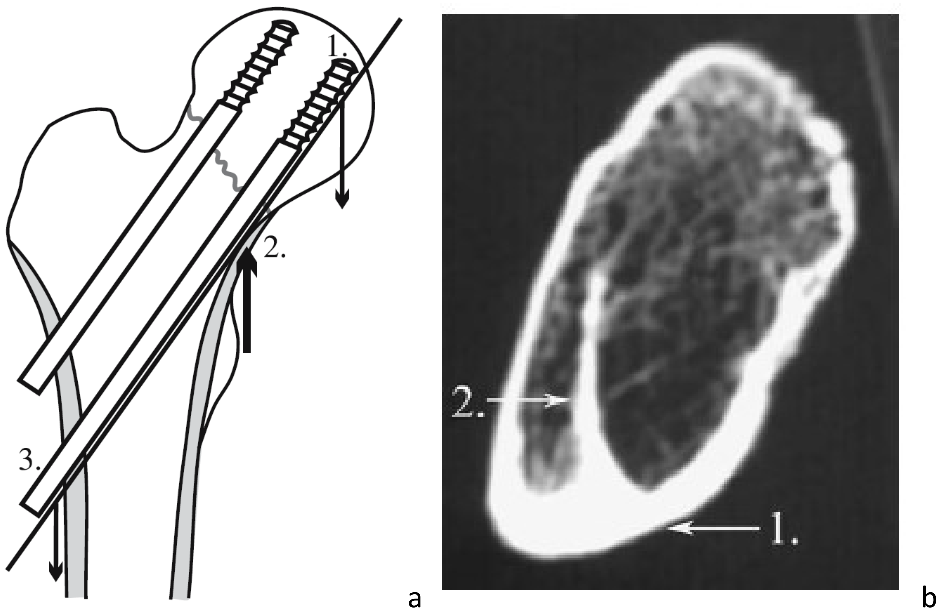


Figure 1.2: (a) Principle of three-point buttressing that is essential in the internal fixation of femoral neck fractures. The subchondral bone of the femoral head (1a), gutter between the Adam's Arch and calcar femorale (2a) and lateral cortex (3) offer mechanical support to the internal fixation device. (b) Lateral radiograph (approx. at the level of the fracture line shown in a) of the femoral neck showing the Adam's Arch (1b) and calcar femorale (2b). Reproduced with permission from Springer (licence number: 4450940728208) [9].

Etiology and fracture patterns

Hip fractures can result from high- and low-energy trauma and are classified by the location, stability, fracture line type, angulation and/or displacement of fractures. While various classification systems exist, none have been adopted universally and there is still debate on which is most ideal. Among the systems, those most commonly used are Garden's, Pauwels, and AO. Some have suggested that Pauwels be used for young and adult patients, whereas Garden's be used for the elderly [12]. Figure 1.3 illustrates the anatomy of the hip and common femoral neck fracture patterns. These patterns can be classified further, as explained below.

First, according to their location, these fractures can be grouped into two main categories: cervical (also known intracapsular) and trochanteric fractures (also known as extracapsular or pertrochanteric). Cervical fractures can be either femoral neck fractures (FNF) or femoral head fractures. Another way of classifying based on location according to whether the fracture is inside the hip joint (i.e. articular); FNF and trochanteric fractures are extraarticular and FHF are articular. FNF can be further broken down to basicervical, transcervical or subcapital (see figure 1.3). Most FNF are subcapital or transcervical [9]. These fractures can have variable fracture line types: smooth, jagged, simple, or comminuted, which highly influence stability and ultimately the type of treatment needed [9].

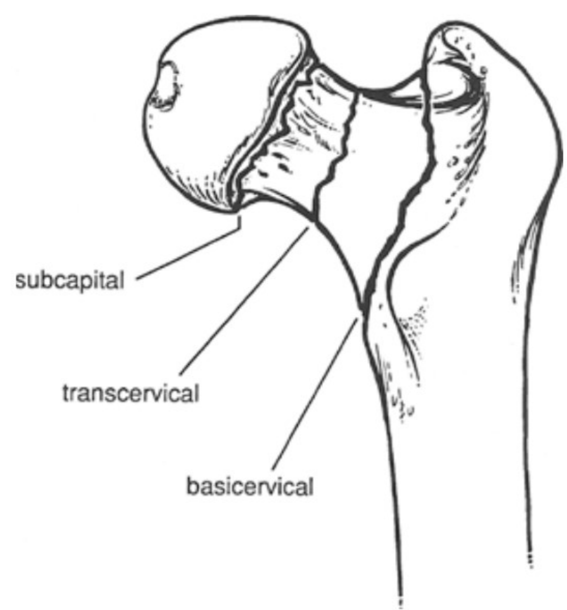


Figure 1.3: Types of femoral neck fractures according to location. Reproduced with permission from Springer (licence number: 4450960195677); [14].

The Pauwels classification system (see figure 1.4) was developed in 1951 and primarily describes the FNF stability and angulation of the fracture line in the anterior posterior plane. According to this system, stable fractures with valgus impaction of the femoral head and angulation under 30̊ are known are Type I. Valgus impaction refers to bone impaction that causes an external rotation of a limb, away from the midline, which is the opposite of varus impaction, when the limb is rotated medially. While this type may begin as stable, bone turnover can lead to loosening to produce an unstable fracture. Those with an oblique fracture line in the anterior posterior plane with angulation of the plane of fracture between 30̊ and 50̊ are Type II and have some support due to the obliqueness of the fracture. Type III have an angulation of the fracture plane up to above 50̊, which often affect the junction between the femoral head cartilage and femoral neck, where the epiphyseal vessels are situated. This type is often very instable and difficult to reduce. While this system is informative, it ignores the angulation of the fracture in the axial plane, and since FNF can be oblique along the axial line, it has a limited classification ability and overall utility [13].

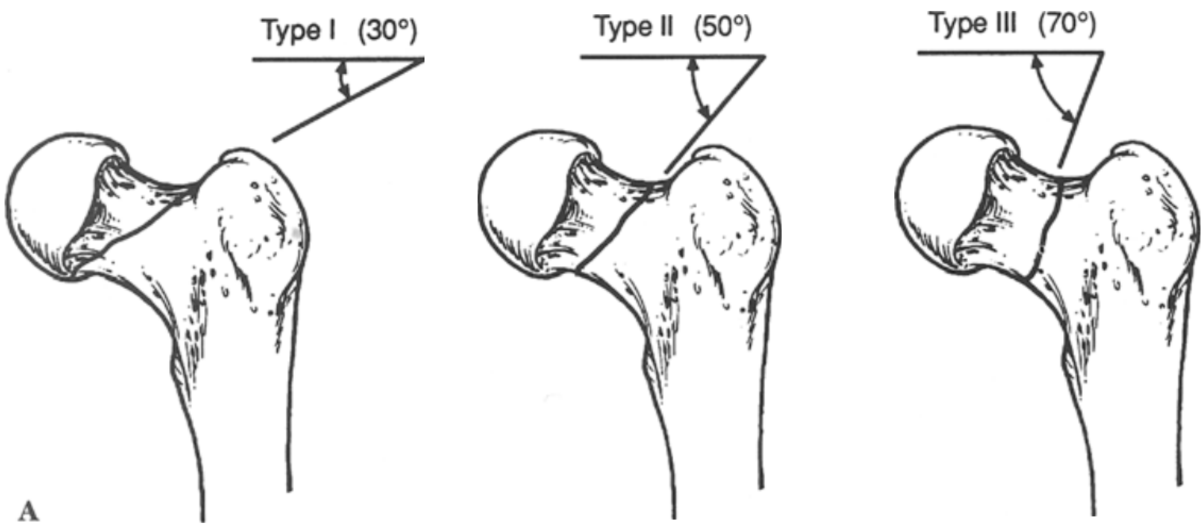


Figure 1.4: Pauwel's classification of hip fractures. Fractures with an angulation below 30̊ are type I, between 30̊ and 50̊ are type II and above 50̊ are type III. (Reproduced with permission from Springer (licence number: 4450960195677); [14].

As mentioned earlier, FNF can also be grouped according to their level of displacement. They can thus be grouped simply as either nondisplaced and displaced, or further broken down using a specific classification system. For example, the Garden's classification (hereafter used in other sections of this text) groups fractures into Types I-IV (see figure 1.5). Types I and II are known as non-displaced and types III and IV are displaced. Type I is used for incomplete fractures with impacted valgus, type II corresponds to complete fractures without displacement, type III is for complete fractures with partial displacement and type IV for complete fractures with full displacement. The latter is recognized as the worst as it is associated with injured retinacula of Weitbrecht, which as described earlier lie on the surface of the femoral neck and house arteries that are critical for the supply of blood to the femoral head; type IV therefore has the highest risk of avascular necrosis, in addition to mechanical complications to instability of the fracture [13].

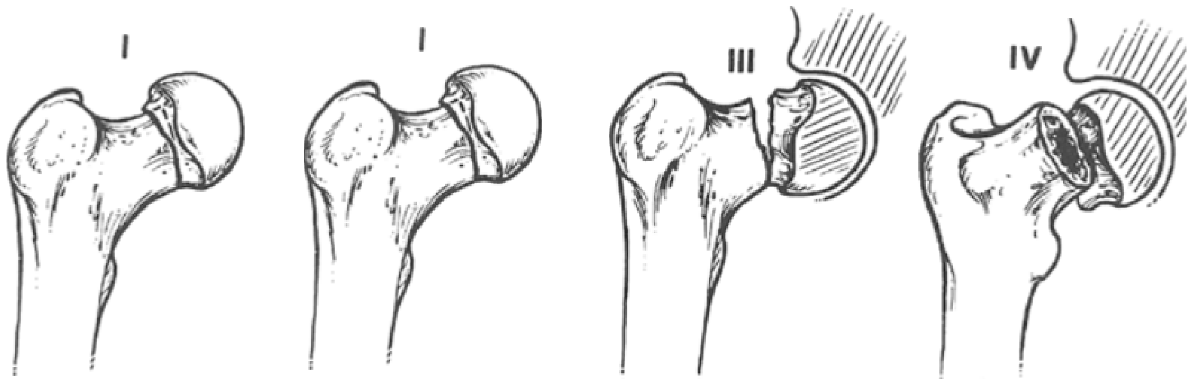


Figure 1.5: Garden classification of femoral neck fractures. Reproduced with permission from Springer (licence number: 4450960195677); [14].

The system developed by the AO combines location and displacement to give three groups (figure 1.6). Type B1 is a nondisplaced subcapital fracture, B2 is a transcervical or basicervical fracture, and B3 is a displaced subcapital fracture. According to an international survey that queried 298 surgeons, Garden's system was the most commonly used. While 72% of respondents preferred it, only 39% judged themselves able to differentiate between each type, yet 96% could differentiate between displaced (Types I and II) and nondisplaced (III and IV) fractures [15].

Figure 1.6: AO classification of femoral neck fractures. Type B1 is a nondisplaced subcapital fracture, B2 is a transcervical or basicervical fracture, and B3 is a displaced subcapital fracture. Reproduced with permission from Springer (licence number: 4450960195677); [14].

One weakness of using the Garden's classification alone is that it relies solely on the anteroposterior radiographs. Some fractures initially found to be type I are then reclassified as type III following lateral imaging [9]. Furthermore, some fractures classified as type III are found to be type IV following lateral imaging. To address this issue, Garden's Alignment-Index can be used as they also consider lateral imaging. To summarize, it evaluates the angulation of the femoral neck in the AP and lateral images to determine if there is displacement. It can be used preoperatively as well as intra- and postoperatively to ensure proper reduction [9].

While high-energy fractures are caused by high-energy impacts such as those from motor vehicle accidents, falls from above standing height, or gunshot wounds, low-energy implies an impact similar to a fall from standing height[16]. Most femoral neck fractures are due to low-energy trauma, such as falling from standing height or twisting of the body with a planted foot [12]. Elderly individuals suffering a HF due to low-energy trauma are more likely to experience a subcapital fracture [12]. Some individuals with such severe bone loss suffer a

FNF without falling; these are known as insufficiency or spontaneous fractures [9]. It has been recommended that these terms be avoided since under investigation, they are often found to be either stress (also known as fatigue) or pathologic fractures. The former occur due to strenuous activity without an identifiable trauma, whereas the latter are caused by bone diseases without an identifiable trauma.

High-trauma accidents on the other hand can result in a variety of patterns, of which femoral neck and ipsilateral fractures (multiple fracture on same side of the body – ex. ipsilateral femoral neck and subtrochanteric fractures on the same leg) are among the most severe[17], [18], [19]. Young adults will more likely suffer a basicervical or vertical distal neck fracture caused by high-energy impacts as seen when an abducted knee received an axial load, as in automobile accidents or high falls[9].

While it may be easier to discuss HF as a whole, the vast difference in the biology and mechanics, and ultimately outcomes between hip fracture patterns and types highlights the need to treat each as their own pathology needing different therapeutic approaches[20].

Diagnosis and Treatment

This section will give a summarized overview of the essential FNF inpatient care, covering diagnosis, pain management, prophylactic measures, and treatments. As there is a substantial amount of research that has been conducted to identify the best practices for FNF patient care, various guidelines have been produced that serve to guide physicians through this enormous amount of research. The guidelines discussed include that from the American Academy of Orthopaedic Surgeons (AAOS), National Institute for Health and Care Excellence

(UK), Bone and Joint Decade (Canada), and the Health Quality Ontario & Ministry of Health and Long-Term Care. Various other guidelines exist, but the reference to guidelines was limited for relevance sake.

Patients with HF often present after a trauma and with hip pain and inability to bear weight, with potential shortening and external rotation of the affected leg. Anterioposterior (AP) and lateral X-rays of the hip are used to confirm diagnosis[21]. With 90-95% sensitivity, the radiograph may not indicate a fracture, in which case the fracture is known to be occult [22, 23]. These cases can also not be associated with physical deformity and inability to bear weight, but only vague pain in the buttocks, knees, thighs, groin or back [23]. In such cases, MRI is then the gold standard for further imaging, offering 100% sensitivity and specificity for diagnosing occult hip fractures [22-24]. Bone scans and CT scanning may be more feasible due to availability and contraindication of MRIs in certain individuals [22-24]. However, the US guidelines recommend against the use of CT scanning due to the lack of quality studies and potential harm caused by radiation exposure [23].

Management of pain is a critical first step in treating HF patients; administration of analgesics can preoperatively facilitate the diagnosis and treatment of other injuries and postoperatively enable mobilization of the injured limb, which ultimately ameliorates overall prognosis. While not much research has been conducted on the best practices for analgesics for HF patients, many national guideline groups have internally come to the consensus that unless contraindicated, paracetamol can be routinely used, and can be supplemented with opioids [24]. They also recommend that non-steroidal anti-inflammatories be avoided.

Regional anaesthesia can be used pre- and intraoperatively to offer adequate anesthesia during surgery, but also to reduce the need for opioids, which can cause delirium [22]. In any case, these methods should not be used to substitute early surgery [24]. The Canadian hip fracture treatment guideline recommends multimodal analgesia as it provides improved pain relief and less side effects [25].

A meta-analysis comparing the two, including twenty retrospective observational and three randomized controlled trials, found that the regional anesthesia results in less in-hospital mortality, myocardial infarctions and respiratory failures, and shorter hospital stays than general anesthesia [26]. However, both had equal 30-day mortality. Several other systematic reviews and meta-analysis have been conducted, each with their own weaknesses though, albeit agreeing with these results. The review mentioned above is among the most recent and offers an in-depth discussion reviewing these other studies [26].

Surgical treatments used vary between patients depending on several factors such as age and fracture pattern, and have been shown to heavily affect outcomes [20]. For most patients, surgical management is considered as the optimal and most cost-effective approach as the benefits of facilitating early mobilization through surgical treatment outweigh the risks of surgery [22]. Nonetheless, some nondisplaced intracapsular fractures can be successfully treated nonoperatively. Some have recommended extended bedrest for nondisplaced fractures if the patient is under the age of 50 years [20]. Others have recommended slight weight bearing with the aid of crutches for six weeks if there has been an excessive delay since fracture or if the presence of comorbidities/frailty of the patient inhibits surgery [23].

Surgery

Once the decision has been made to operate, delays should be avoided as evidence suggests that delays contribute to poor outcomes. While exact optimal timing remains controversial, consensus can be interpreted as the following: surgery should be undertaken as soon as the patient's medical condition permits. Some experts recommend that surgery be conducted within 24h of hospital admission to reduce the rate of complications and mortality [23], and the US, Canadian and NICE guidelines states that surgery should be conducted within 24 to 48 hours [24, 25, 27]. However, since HF patients are often elderly with comorbidities (heart failure, chronic obstructive pulmonary disease, and diabetes mellitus, among others), some suggest that optimization of the patient's medical status can permit some delays in surgery, with a maximal delay of 72h [23]. Early surgery is argued to be favored as reduction of the fracture optimizes positioning of the blood vessels in the femoral neck and decreases intraosseous intracapsular pressure [28].

In addition to anesthesia/analgesia, preoperative preparation includes antibiotic and thromboprophylaxis. Antibiotics should be administered within 2 hours of the start of surgery, and 8 and 16 hours postoperatively. Cefazolin has been accepted as an effective antibiotic, however vancomycin can be use for patients that are allergic to penicillin, or in environments with a high rate of methicillin-resistant *staphylococcus aureus* cases [23]. Regarding thromboprophylaxis, various drugs and mechanical approaches have been evaluated. The Canadian guideline recommends pharmaceutical thromboprophylaxis only if the delay to surgery exceeds 24 hours. Fondaparinux has been shown to have 56.4% lower risk of venous thromboembolism than enoxaparin, a low molecular weight heparin, but is more difficult to use preoperatively as it has an 18-hour-long half-life and could interfere with anaesthesia and/or cause excessive intraoperative bleeding. Other common low molecular weight heparins available are Tinzaparin and Dalteparin. There is also low-dose unfractionated heparin available, which when compared to low molecular weight heparin, offers similar thromboprophylaxis. Individual patient characteristics should guide the selection between the two. Warfarin has also been compared to other drugs, with mixed results. Yet, it has been shown to be more effective than aspirin, which has been recommended to not be used alone for prophylaxis. Mechanical prophylaxis consists of graduated compression stockings and intermittent pneumatic compression devices and has been shown to reduce the rate of DVT when compared to placebo. Yet, it is recommended that it be used only in patients at high risk, and in combination with pharmaceuticals [22]. However, the Canadian guideline does not recommend mechanical prophylaxis.

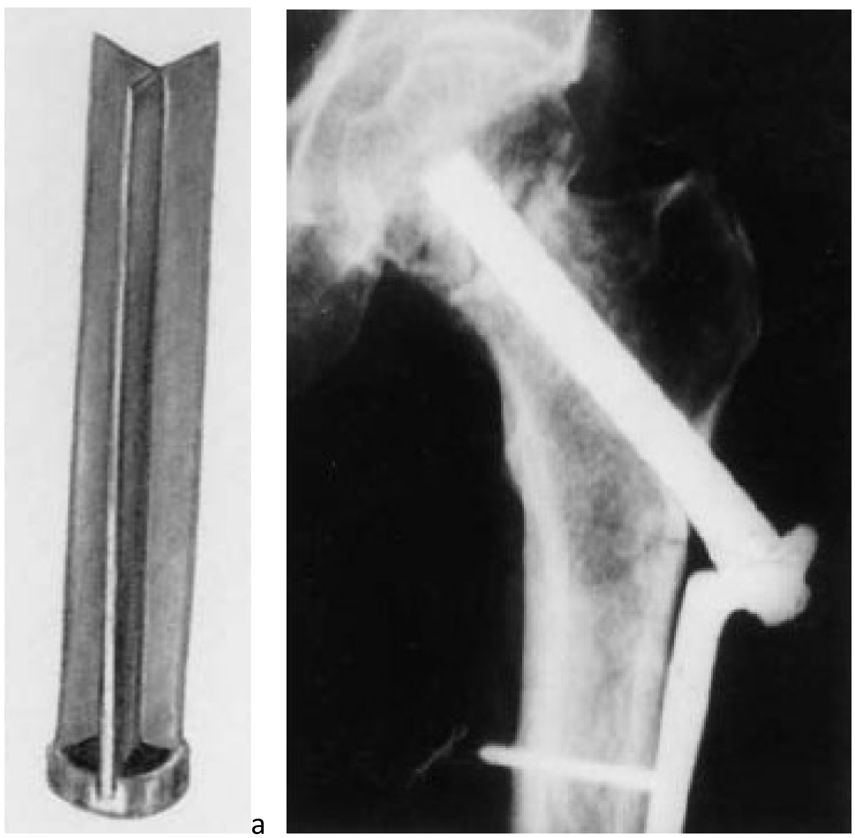
With the purpose of stabilizing, aligning and immobilizing hip fractures, surgical orthopaedic intervention permits proper anatomical healing and minimizes the risk of complications related to prolonged bed rest [20]. Surgeons should aim for the patient to be able to weight bear immediately after surgery [24]. Various approaches are available, and selection is dependent on various factors. Some have been found to be superior, but there remains a lack of consensus of which is best for displaced femoral neck fracture in the elderly. Hence, many authors still refer to FNF as the “unsolved fractures”, even though it had been given this title in 1953 [29-31]. Some go as far as stating that it "is one of the most difficult problems to orthopaedic surgeons all over the world" [29].

Reduction and Internal Fixation

Reduction of the fracture can be attempted once adequate analgesia/anesthesia has been administered. Closed reduction should be attempted prior to open reduction to minimize invasiveness, however if needed (ex. in displaced fractures that cannot be reduced with a closed approach), open reduction should be conducted. Closed reduction can be achieved with slight traction and internal rotation x-ray overview. The Leadbetter maneuver can also be used. Closed reduction should not be forced; if not easily achieved, open reduction should be used [32].

While over 100 different internal fixation implants could be counted in 1974 [33], common modern internal fixation techniques can now be summarized to include the following: cannulated screws, hook-pins, nails, intramedullary-neck screw devices, dynamic hip screws or sliding nail plates. A brief review of the development of devices is presented here as it gives a logical flow of improvements from one device to another and highlights advantages and disadvantages. To summarize IF altogether, IF has the advantage of faster and easier to perform surgery, saves the natural joint, however it is associated with high failure and reoperation rate [34].

The first major advancement in the treatment of hip fractures was the development of the Smith-Petersen three-flanged nail in 1925 (see figure 1.7a). The flanges on this nail prevented rotation. This was soon replaced by a cannulated nail developed by Johansen in 1932. The mechanical instability of this type of device led to the addition of a plate at the distal end of the nail (see figure 1.7b) to stabilize it along the femoral shaft (as seen in the Thornton Nail). In the elderly, these devices often perforated the femoral head as the femoral neck shortened due to device impaction. In young patients, the nails tend to become distracted and fracture due to fatigue. Furthermore, the need for nailing of these implants is a difficult procedure and results in trauma and can affect femoral head vitality [9].



a b

Figure 1.7: Three flanged nail device similar to that introduced by Smith-Peteren (a) and radiograph of implanted nail with endplate inserted into lateral cortex as in the Thornton nail (b). Reproduced with permission from Springer (licence number: 4450941137684); [9].

Screws were introduced to the treatment of HF in 1951 by Putti [9]. They have since been modified to have a hollow shaft, hence why they are called cannulated screws (see figure 1.8). Their hollow cavity allows them to be guided into the bone using guide wires, which facilitates proper screw placement. As with the other techniques reviewed, radiographs are used to ensure appropriate alignment. Screws have tended may be favored over other fixation methods in some settings, which may be due to its lower invasiveness, preservation of cancellous bone compared to larger devices, improved rotational stability thanks to the installation of more than one device [35], and better anchoring of screws into subchondral due to the threading of the ends of the screws. However, insertion of multiple screws is difficult and improper installation can lead to disrupted impaction.

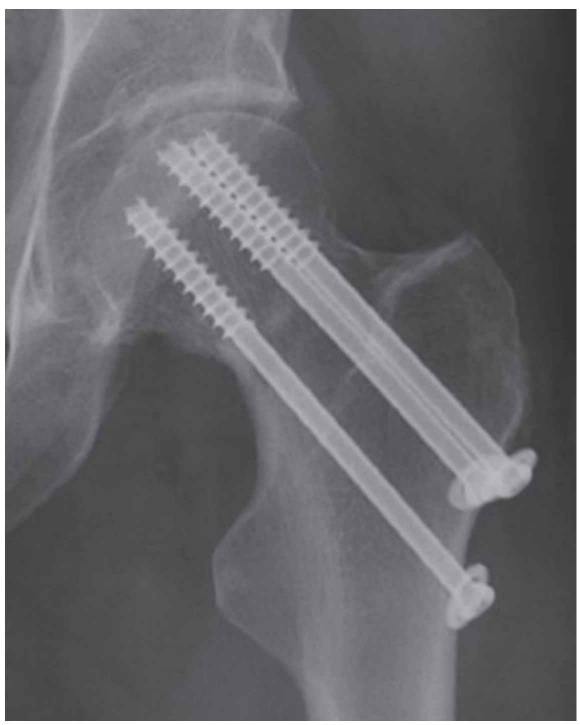
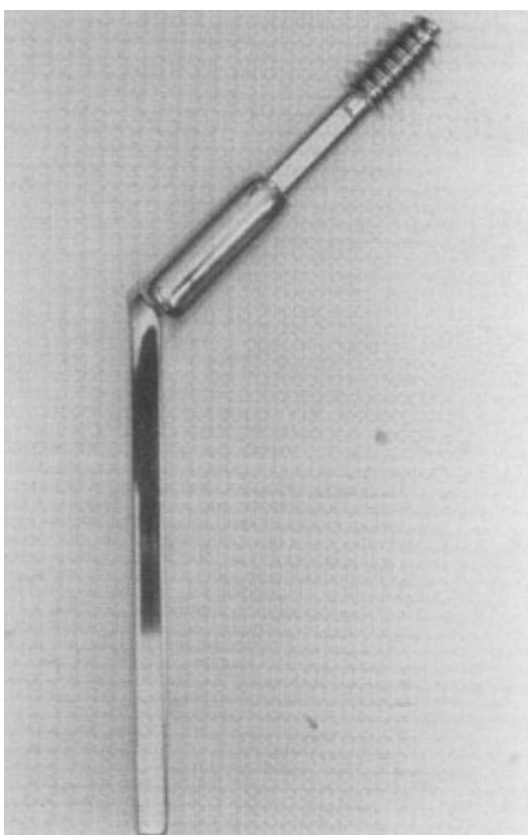


Figure 1.8: Radiograph of cannulated screws. Reproduced with permission from Springer (licence number: 4450960195677); [10].

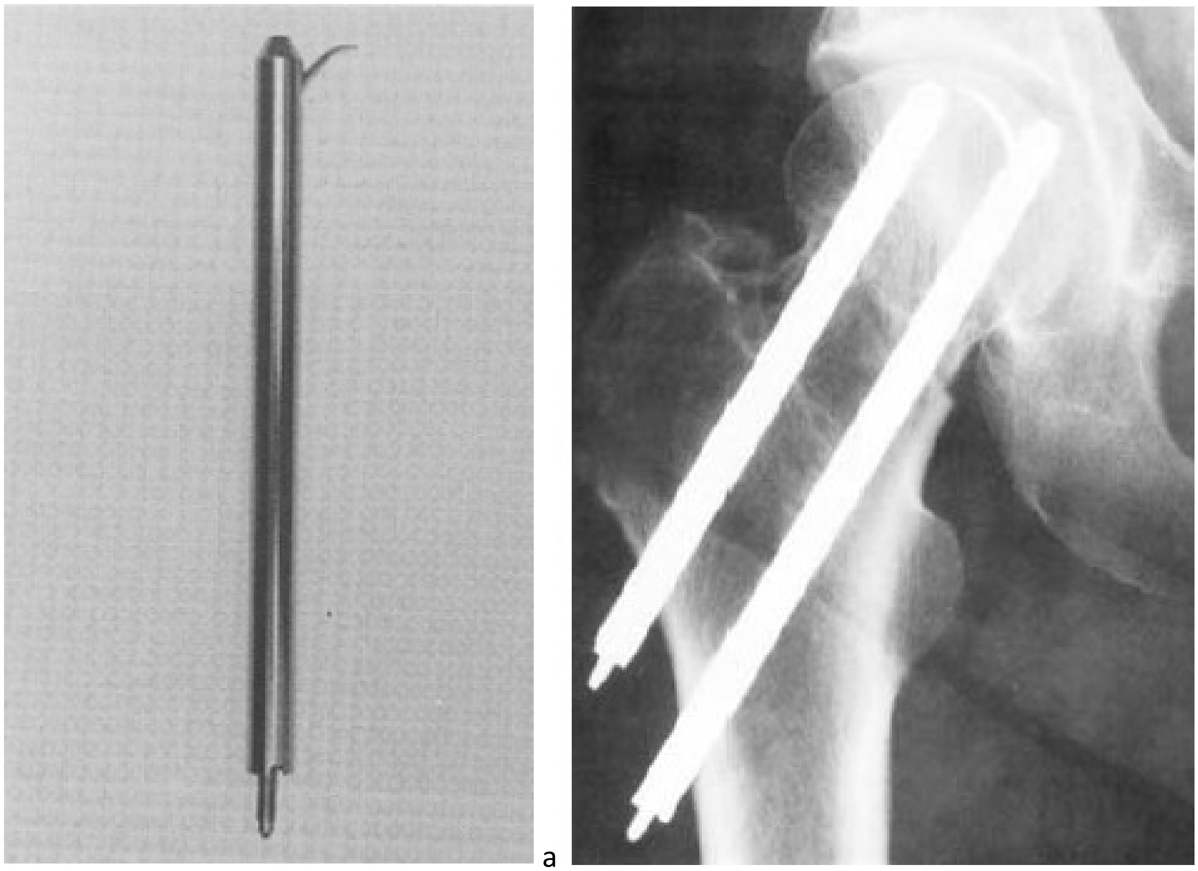
To compensate for the settling of the fracture, a dynamic component was added to the screw system to give the dynamic hip screw (DHS, see figure 1.9) [36]. The femoral component of the DHS is screwed into the lateral cortex, stabilizing it. The sliding nail plate (SNP) makes use of a similar mechanism, however with a nail instead of a lag screw. These devices are argued to provide better mechanical strength as determined from biomechanical studies on cadavers [35, 36]. The disadvantage of these devices is that they do not restrict rotation of the femoral head. To overcome this, the hole in the plate through which the screw is inserted and the distal end of the screw were changed to be square. This was found to not entirely prevent rotation and an additional screw was added to the dynamic devices to overcome this [9].


a

b

Figure 1.9: Photograph of a dynamic hip screw (a) and radiograph of implanted dynamic hip screw with an additional screw to help resist femoral head rotation (b) Reproduced with permission from Springer (licence number: a. 4450960195677, b. 4450941404121); (a. [14], b[9]).

Hook-pins (also known as the Hansson hook-pins, see figure 1.10) were first developed for the fixation of slipped capital femoral epiphysis in children and has since been used in femoral neck fracture fixation since the 1970s, but only first described in the literature in 1983. As with cannulated screws, a guide wire is inserted first to facilitate alignment. This wire is then used to guide a drill, which will vacate enough space to easily fit the pins in. The holes are made to be slightly bigger than the nails to minimize trauma to the femur during insertion. As with cannulated screws, it resists rotation of the femoral head. The distal pin or screw prevents valgus angulation, while the proximal pin or screw prevents dorsal angulation. Parallel placement of the pins allows for compression of the femoral head to occur, and their small size minimizes bone disruption. The small incision needed results in lower trauma and quicker healing times.



a b

Figure 1.10: Photograph of Hansson hook pin (a) and radiograph of Hansson hook pins for a femoral neck fracture. Reproduced with permission from Springer (licence number: a. 4450960195677, b. 4450941137684); (a.[14], b.[9]).

Various systematic reviews and meta-analyses have been conducted to determine which IF implant is best, but none have produced conclusive results that support one over another. Furthermore, cement or bone grafts have been investigated for their use in internal fixation surgery to improve outcomes.

Arthroplasty:

Historically, this procedure has undergone an enormous amount of innovation, with the first being attempted in 1891 to replace hips damaged by tuberculosis with ivory in lieu of a femoral head [37]. Followed by attempts involving transplantation of various tissues, including xenographs of pig bladder submucosa, hip arthroplasty surgery continued to be improved through the ages [37].

Modern arthroplasty can be categorized as either unipolar hemiarthroplasty, bipolar hemiarthroplasty or total hip arthroplasty. Hemiarthroplasty refers to the replacement of the proximal femur with an implant, with either a single component (unipolar, see Figure 1.11) or dual component articulated implant (bipolar, see figure 1.12). Bipolar hip arthroplasty has the advantage of reducing the risk of acetabular wear in those with life expectancy over five years and is therefore best for those under 80y. Its disadvantage is that is it more expensive, can have polyethylene wear that can lead to mechanical loosening, and dislocation and may require open reduction in some cases [38]. Unipolar on the other hand has the advantage of not having material wear that can contribute to eventual mechanical loosening of the implant, but the disadvantage of being associated with acetabular wear [39].


a


b

Figure 1.11: Photograph (a) and radiograph of implanted (b) unipolar hip hemiarthroplasty. Reproduced with permission from Springer (licence number: 4450960195677); [14].

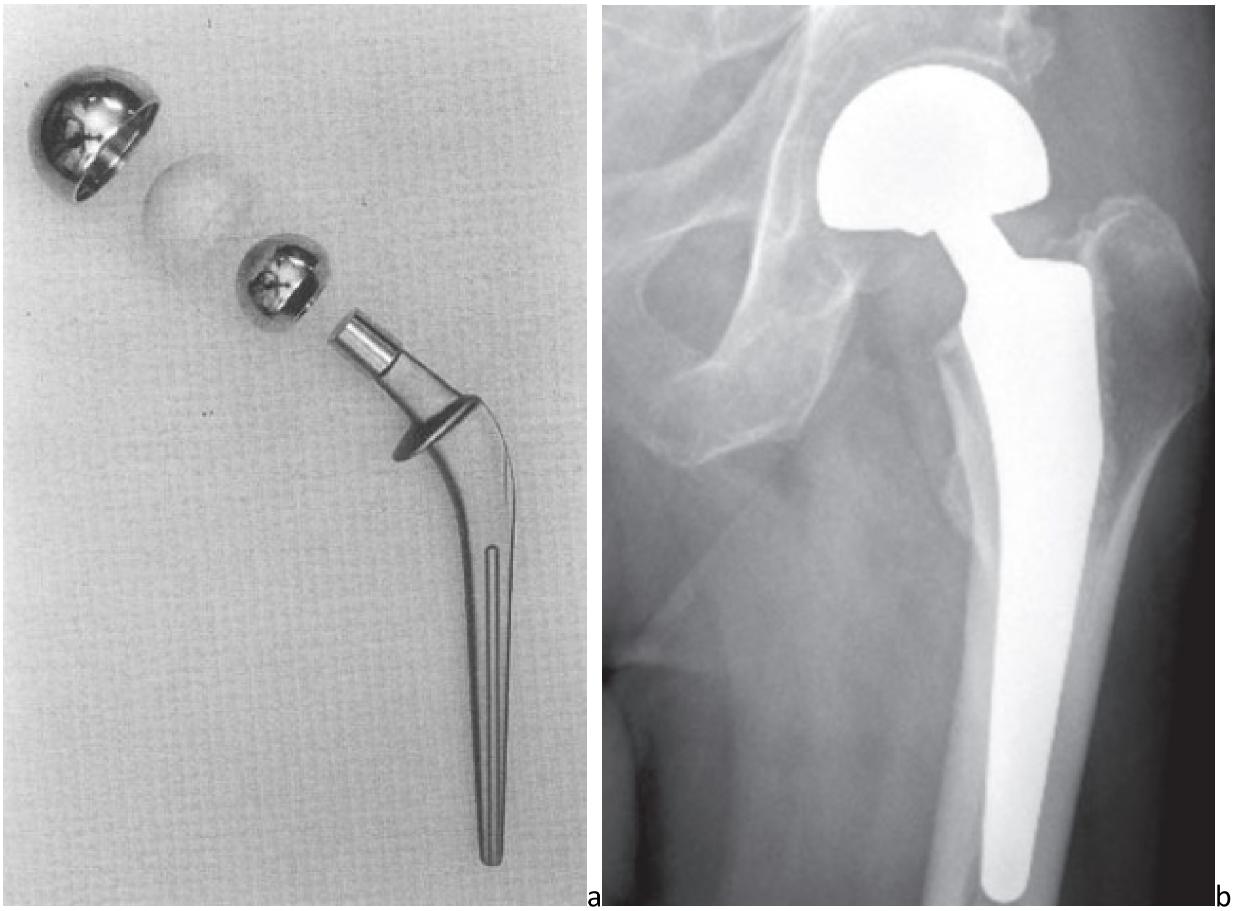


Figure 1.12: Photograph of a bipolar hip hemiarthroplasty (a) and radiograph of implanted bipolar hip hemiarthroplasty. Reproduced with permission from Springer (licence number: 4450960195677), [14].

Total hip arthroplasty refers to the replacement of both the proximal femur and acetabulum (see figure 1.13). It has the advantage of high predictability, low pain and high overall function, no need for revision surgery due to failed IF or acetabular wear of HA, but the disadvantages of increased cost, surgical time, and blood loss, greater dislocation that HA and IF [38]. Various materials can be used for these implants, combinations include metal-on-polyethylene, ceramic-on-polyethylene, ceramic-on-ceramic, metal-on-metal and ceramic-on-metal [40]. In addition to the materials used, several femoral component designs exist. For example, the femoral head size can vary between implants (22.2mm to 50mm in diameter). They each have advantages and disadvantages and the selection can depend on the patient's needs. A network meta-analysis comparing these combinations for primary total hip replacement found that metal-on-metal, small-head implants led to an increased revision risk, but there was no significant difference in Harris Hip Scores [40]. Femoral components (i.e. stems) can also vary. Various designs have been used over the past few decades, and some have proven to be more successful than others. The NICE guidelines recommend against the use of the Austin Moore and Thomson stems, and prefer those with Orthopaedic Data Evaluation Panel ratings of 10A, 10B, 10C, 7A, 7B, 5A, 5B, 3A, or 3B [27].

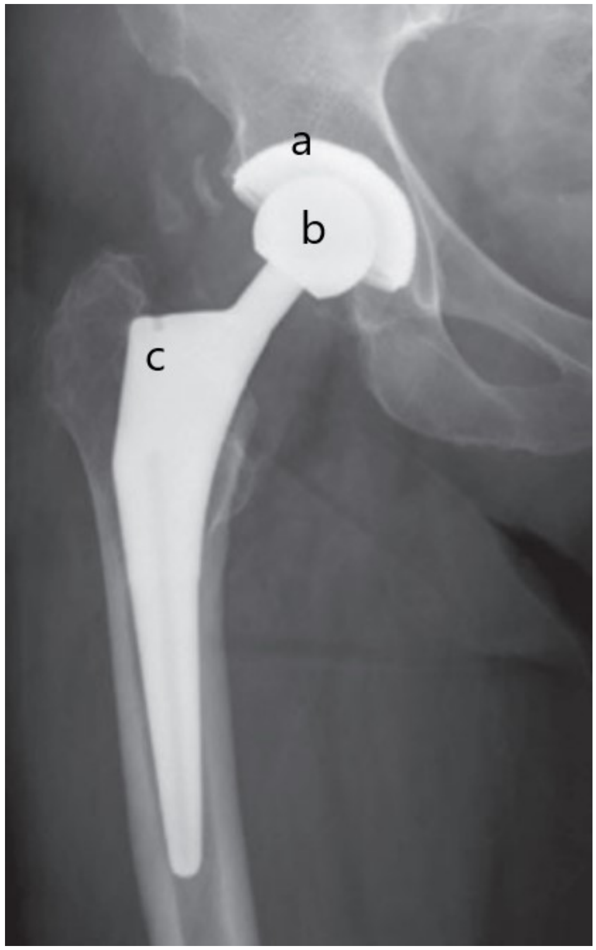


Figure 1.13: Total hip arthroplasty for a femoral neck fracture. The device is composed of three parts, the acetabular shell (a), femoral head (b) and femoral shaft (c). Reproduced with permission from Springer (licence number: 4450960195677); [14].

Arthroplasties can be supplemented with cement. They can be cemented, uncemented, hybrid or reverse hybrid; cementing of both the femoral and acetabular components is cemented, cementing of the femoral but not acetabular component is known as hybrid, whereas cementing of the acetabular but not the femoral component is known as reverse hybrid. The NICE and AAOS guidelines recommend the use of cement for arthroplasties [27].

There are also various surgical approaches that can be used for hip arthroplasties. To summarize, posterior (also known as southern, Moore, and dorsal approach), direct lateral (also known as anterolateral, transgluteal and Modified Hardinge approach) and direct anterior (also known as modified Hueter approach) are most common [41, 42]. Each has advantages and disadvantages, however the AAOS and NICE guidelines recommend the direct lateral approach [24, 27]. The AAOS refers to two studies as their rationale that show lower rates of dislocation [27, 43, 44]. While these studies were non-randomized and retrospective, evidence nonetheless shows that there is a difference in outcomes for each approach. The Ontario guideline recommends the approach the surgeon is most used to.

For the management of young patients, the goal of surgery is to preserve the femoral head and arthroplasty is therefore disfavored[45], [46]. Preservation of the femoral head should be favored even in cases in which there is substantial delay to surgery. A retrospective study investigating young patients with an average of 13 days delay to surgery (range of 2 to 30 days) demonstrated that all patients had complete union and only two suffered from AVN (non-union).

While internal fixation may be the most successful approach for minimally and nondisplaced intracapsular fractures, the ideal surgical management of displaced intracapsular fractures is not as well understood. Several RCTs and meta-analyses have been conducted to elucidate the matter, and guidelines have made various recommendations based on this research. Here is a review of the evidence available. While FNF may be treated with ORIF, many require arthroplasty. In fact, the AAOS, Ontario and NICE guidelines recommend that arthroplasty be used for displaced FNF over IF [24, 27]. Furthermore, the Ontario Hip Fracture Guidelines recommend total or hemiarthroplasty for displaced FNF in patients over 65 years, and that total hip arthroplasty be used for younger, more cognitively healthy individuals. Alternatively, the NICE guidelines recommend that THA be used instead of HA for patients that were ambulatory with the aid of no more than a walking stick, are not cognitively impaired, and medically fit for the surgery [24]. The basis for recommending THA over HA for younger or more fit/ambulatory patients is that HA is associated with acetabular wear in active individuals. These recommendations are based on several studies comparing ORIF to THA, which found that THA led to lower rates of reoperation, pain, functional status and complication rates. Reoperation was reported consistently as a more common complication in IF than THA. Nonetheless, there remains a lack of strong evidence supporting THA over HA.

Burden of illness

To identify the most optimal treatment for FNF, there is a need to evaluate the dangers of each treatment and compare them. The outcomes that can occur following a FNF will therefore be enumerated, and those that are most frequent and severe for each treatment will be identified. Risk factors that increase the risk of some of these outcomes will also be discussed. While FNF are associated with a wide range of complications, this study will focus on mortality, quality of life (QoL), and avascular necrosis of the femoral head (AVN) as the complications of interest. While all treatments can be compared for the former two, AVN can only occur following IF, and individual IF options will therefore be compared in this study. Only outcomes that will either directly be analyzed in this study or those that contribute to the risk of these outcomes will be discussed.

Even with advances in FNF patient care, these injuries still have the potential of death, pain, poor healing, substantial functional disability, and decreased patient independence for managing daily activities [20] [47]. FNF result in an extremely wide range of outcomes, depending not only on the severity of the injury, but also on the fracture characteristics, treatment used and several other variables such as age and comorbidities.

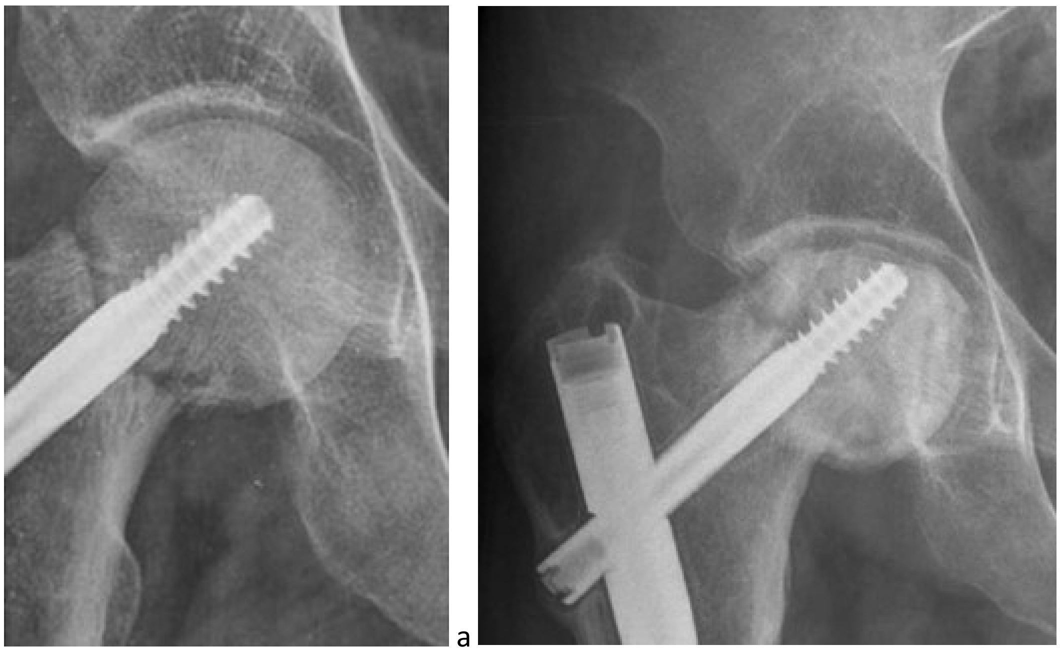
The most severe outcome that can result from FNF is death and 12-20% of Canadians suffering a FNF will die in the first year [48]. In fact, osteoporotic fractures kill more women than myocardial infarctions, coronary heart disease and breast cancer in the United States in one year, and FNF constitute a substantial portion of these fractures [49]. Results from the literature vary but the majority state that the risks of death and complications are higher in men than women and increases with age [50][51]. Furthermore, comorbidities have been found to be predictors of postoperative morbidity and mortality in this population [51].

Various studies have investigated mortality risk factors for FNF. Individual comorbidities as well as the combination of comorbidities, such as cardiac, psychiatric and pulmonary diseases, have been found to be significant predictors of death [52, 53]. Comobridity/health indexes have also been found to be predictors; these include the Charlson Comorbidity Index, ASA, Hip-Multidimensional Frailty Score, Nottingham Hip Fracture score, SERNBO score [50-52].

In addition to comorbidities, the occurrence of complications also increases the risk of death. To determine how complications contributes to this risk, a retrospective analysis of 8930 patients in the United States investigated the effect of complications on the occurrence of complications [53]. Among their main results, they found that having any complication led to a significantly higher mortality rate at 30 days and 1yr. Cardiac, pulmonary and cerebrovascular complications were associated with the highest mortality rates; both, or combined, cardiac and pulmonary complications we associated with a significantly greater risk of death than either of them individually.

While most patients will not suffer any negative outcome, approximately one fifth of surviving hip fracture patients experience medical complications [53]. These include cognitive and neurological, cardiac and/or vascular, pulmonary, gastrointestinal, urinary tract, hematologic, endocrino-metabolic, and other problems. More specifically, infection, peri-implant femoral fractures, prosthetic dislocation, acetabular protusion, implant loosening, decubitus ulcers, fat emboli, deep venous thrombosis, pulmonary emboli, pneumonia, acute urinary retention, myocardial infarctions, and strokes have been associated to FNF [17, 54] [55]. While some are specific for a certain treatment (ex. AVN cannot occur following THA), others can occur in any patient following FNF.

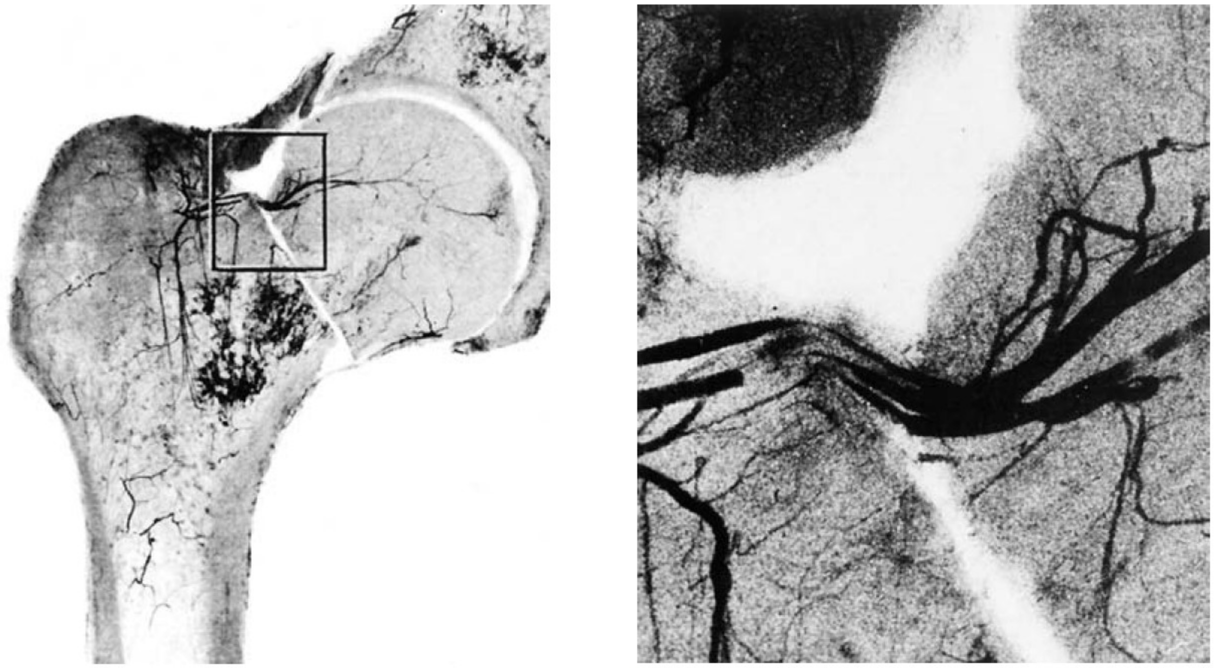
The main problems in FNF are caused by the lack of periosteum (the lining around the bone, that facilitates fracture healing among other functions), proximity to a joint, and obstruction of blood flow to the femoral neck [56]. Treatment of FNF with IF methods can thus lead to various problems, two major ones being non-union and avascular necrosis of the femoral head (AVN – see figure 1.14).



a b

Figure 1.14: Radiographs of non-union (a) and avascular necrosis of the femoral head (b) in patients with femoral neck fractures. Note the fracture line that remains in the case of non-union, and the collapse of the femoral head in the avascular necrosis case. Images reproduced with permission (email correspondence)[56].

Risk factors for AVN include displacement, age, and time to surgery, and this complication occurs in 9-18% of patients. Technically, AVN is due to the obstruction of blood flow to the femoral head. Recall that the medial circumflex artery feeds into the posterocaudal arteries, then into the lateral epiphyseal network, and that this is a critical blood supply to the femoral head. Displacement of the femoral head can disrupt these vessels and obstruct blood flow (see figure 1.15). Studies on mice have shown that delays of as little as 6 hours results in a loss of femoral head vitality [9].

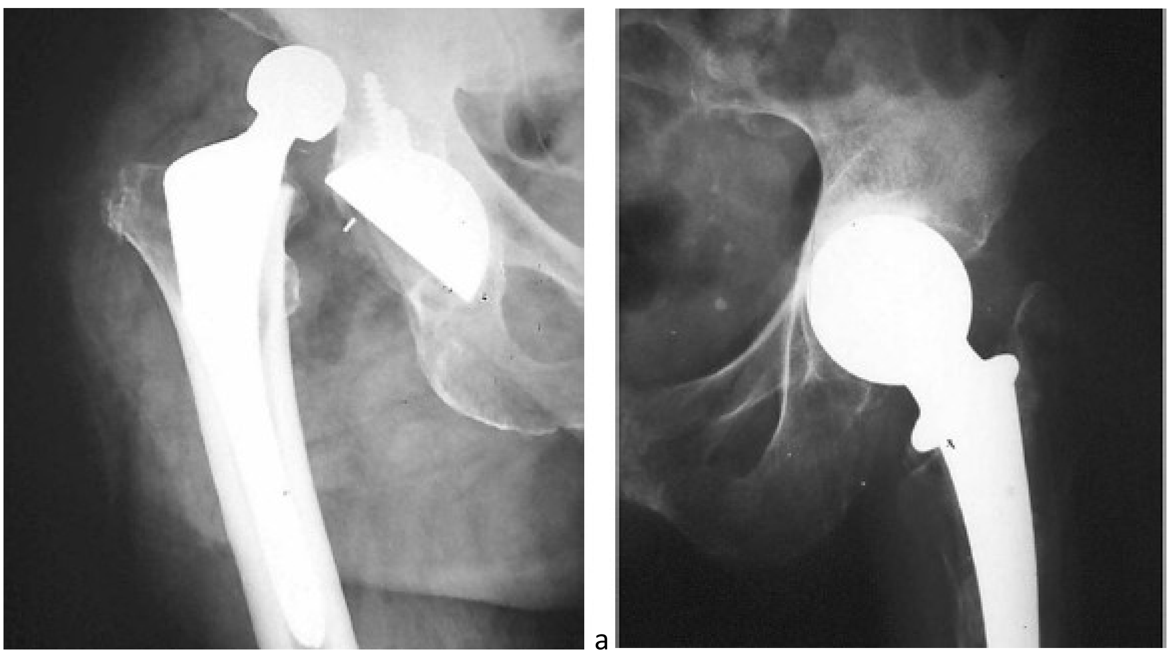


a b

Figure 1.15: Barium angiogram of a cadaver showing kinking of vessels following a femoral neck fracture. Expedition of surgery is optimal to regain blood flow. Overview (a) and four-times magnification (b) are shown here. Reproduced with permission from Springer (licence number: 4450940898567) [9].

As these major complications lead to an unviable femoral head, reoperation is required. Many experts have therefore opted for arthroplasty as evidence shows that there is significantly lower rates of complication and reoperation. AVN risk is also dependent on the specific fracture pattern; basicervical fractures rarely lead to AVN and their treatment differs from transcervical and subcapital fractures [12]. It must therefore be noted that not all FNF have equivalent risk of AVN.

When treated with arthroplasty, FNF are associated with several complications including dislocations, which is most common in THA, and acetabular erosion, which most common in active individuals with a HA (see figure 1.16). To minimize the latter, many recommend that THA be used in these more active patients. Uncemented arthroplasty can often be associated with higher levels of pain and a large risk of later femoral fractures, although they do offer higher hip scores. Logically, as there is no more femoral head, the risk of AVN is null. Alternatively, as the femoral head is replaced in arthroplasty, the risk of device dislocation or acetabular/device erosion is absent in internal fixation.



a b

Figure 1.16: Radiographs of dislocation following total hip arthroplasty (a), and acetabular erosion following hemiarthroplasty (b). Image reproduced with permission (email correspondence).

In addition to evaluating individual outcomes, Quality of Life (QoL) indicators offer a single descriptor of a patients' wellbeing and can be useful for comparative purposes in certain cases. Hip fractures can highly affect an individual's QoL. Salkeld and colleagues (2000) used the time trade off technique to estimate the utility (preference for health) of women over age 75 that are at risk of hip fractures. Of the women surveyed, 80% would prefer a health state equivalent to being dead than to experience the loss of independence and QOL associated to severe hip fractures [57]. There are several indexes that have been developed to measure QoL, including but not limited to the EuroQoL-5D (EQ-5D), Short Form-6D, Health Utility Index and Nottingham Health Profile.

The EQ-5D is a commonly used standardized instrument that was developed in the 80's by a group consisting of experts from five different European Countries. It consists of two parts, one where the patient is asked to evaluate their QoL according to five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and the second asks the individual to summarize their overall state, generally using the visual analogue scale. There are three version of this test, the 3L, 5L and youth; the former two will be of interest for this paper. The 3L asks respondents to rank each dimension on a scale of one to three, and the 5L is from one to five. This development was to avoid the ceiling effect (inability to differentiate between different health states) caused by having only three levels. The direct output of the EQ-5D is a five-digit long code that represents the scores assigned by the patient for each dimension. This code is then converted to a weighted index score that considers preferences for each dimension. Different countries will have varying preferences and therefore, one five-digit code in one country may have a different value score for individuals living in other countries. This instrument is often used to measure the quality-adjusted life years for economic valuation analyses. While several other tools exist, the EQ-5D has several advantages, principally that it is easy to complete. Furthermore, it has been proven to be useful in many fields, including hip fracture treatment [58], and algorithms exist to convert other index scores to EQ-5D [59].

Factors affecting QoL are associated with the surgical treatment and post-operative and post-acute care models. As described above, patients need to be monitored by multidisciplinary team following surgery and their potential medical abnormalities need to be managed appropriately. A systematic review conducted in 2010 evaluating factors affecting QoL has found that evidence supporting specific rehabilitation programs was conflicting. They report that hip replacement should be favored over IF but cannot identify the best hip replacement method. Lastly, they present data on the effect of nutrition on QoL but conclude that data is insufficient. In addition to these risk factors, adherence to the standard of care recommended in guidelines results in a minimization of the risk of death and complications and thus increase the wellbeing of patients.

Rationale

As previously described, HF are associated with a substantial burden of illness and mortality. The incidence of these injuries in rising due to aging populations, and this is predicted to result in an increase in its healthcare burden. Since there are several treatment options available, physicians need to choose among them. To facilitate this decision making, many randomized controlled trials (RCTs) have been conducted to compare treatments head-to-head. Nonetheless, there remains a lack of such comparisons. While this could be solved by conducting more randomized controlled trials (RCTs), these studies require considerable resources and ethical consideration. Fortunately, recent advances in statistics have led to the development of methods that enable the estimation of relative treatment effects, mathematically, and can then give a single ranking of treatments. These methods are known as indirect treatment comparisons, mixed treatment comparisons or network meta-analysis (NMA) [60]. They will hereafter be referred to as NMAs.

Therefore, as HF are serious injuries that have a severe impact on the patient and healthcare system, there is a need to reduce the burden of illness of HF by ranking treatments according to quality of life, mortality and complications. This can be accomplished using NMAs. In addition to enabling the indirect estimation of head-to-head comparisons of treatments, NMAs have the added utility of pooling the results of various studies to find a "true effect". As in meta-analyses, NMAs find summary values for the pairwise contrasts by pooling effect measures according to weights. This analysis therefore increases the sample size of comparisons and the overall power of each comparison, thus improving upon the findings of individual studies.

Research Questions

Before discussing the objectives and hypotheses of this study, it is important to identify the exact research questions this study is aiming to answer. A useful method to accomplish this for meta-analyses is the PICO framework[61], which involves identifying the population of interest (P), the intervention being investigated (I), an intervention that will be used as a comparator (C), and outcome(s) of interest (O)[61]. Since this framework was developed for meta-analyses, which compare only two treatments and an NMA includes more than two treatments, this approach can be slightly modified to allow for multiple comparisons; “I” can simply be a summary term for all treatments that then make up “C”; the research questions, as per the modified PICO framework, can be formulated as shown below:

Table 2.1: Research question design using the PICO framework. For a complete list of treatments included, please refer to the search query.

PICO component	Research question component
Population	Displaced femoral neck fracture patients
Intervention	Surgical femoral neck fracture treatments
Comparators	Every possible surgical treatment
Outcomes	Quality of life (EQ-5D), mortality and AVN

Following this methodology, the research questions addressed in this study are the following:

i. Which surgical treatment is best for optimizing quality of life following a displaced femoral neck fracture?

ii. Which surgical treatment is associated with the lowest risk of mortality in displaced femoral neck fracture patients?

iii. Which surgical treatment is associated with the lowest risk of avascular necrosis of the femoral head in displaced femoral neck fracture patients?

Note that the surgical treatments for FNF fractures for the QoL and mortality analyses will be classified as follows: cemented total hip arthroplasty (THAC), uncemented total hip arthroplasty (THAU), cemented bipolar hip hemiarthroplasty (BPHAC), uncemented bipolar hemiarthroplasty (BPHAU), cemented unipolar hemiarthroplasty (UPHAC), uncemented unipolar hemiarthroplasty (UPHAU), and internal fixation (IF). Treatments for analyses on AVN will be classified as follow: cannulated screws (CS), nail, pins, dynamic hip screws (DHS), sliding nail plate (SNP), cemented cannulated screws (CSC), cannulated screws with fibular graft (CSFG) and cannulated screws with iliac graft (CSIG). Quality of life can be measured using various indexes and several have been used in RCTs comparing hip fracture treatments, including but not limited to the EuroQol-5D (EQ-5D), Short Form-36 and Nottingham Health Profile. Ultimately, the EQ-5D was chosen as it has been validated as an index appropriate for the QoL evaluation of hip fracture patients[58]. It was also the most common index used in RCTs comparing hip fracture treatments, and several algorithms allow for the conversion of other indexes to EQ-5D.

Once a specific index was chosen, the units that would be used in the NMA to compare treatments need to be defined. This is known as the treatment effect measure (also known in the literature as the effect size measure), and several options exist. While the selection of a treatment effect measure for mortality and AVN was relatively straight forward (odds ratios), that for quality of life required some strategizing.

The Mean Difference (MD) and Standardized Mean Difference (SMD) were evaluated as candidates for the EQ-5D treatment effect measure. As shown below, the MD (μ) is simply the difference between the mean response (EQ-5D score) of the experimental group (μ_e) and the mean response of the control group (μ_c).

$$(\mu) = \mu_e - \mu_c$$

On the other hand, the SMD is a dimensionless treatment effect measure that allows for outcomes measured on different scales to be compared. This is accomplished by dividing the MD by the standard deviation of either a single treatment group or both. Various formulae exist to calculate this treatment effect. The one used in this study is *Hedge's* g (g), which as shown below, makes use of the pooled sample variance (s_p^2). Furthermore, a correction factor $\left(1-\frac{3}{4n-9}\right)$ is used to correct for bias in the estimated standard error, where n is the total number of participants in both groups.

$$g = (1 - \frac{3}{4n-9})\frac{MD}{s_p^2}$$

Since the EQ-5D index is a preference-based score, the scores measured in different populations reflects that populations subjective wellbeing. Since preferences are different among countries, there have been concerns about directly comparing EQ-5D scores from different populations. Furthermore, there are two versions of the EQ-5D, the EQ-5D-3L and EQ-5D-5L and comparing populations that have been evaluated using different versions of the EQ-5D also causes concern. Since the papers included were from different population and most papers evaluated did not report which version of the index was used, a decision was made to use the SMD as it would standardize the scores in case of disparity.

While the randomization of patients in RCTs aims to ensure that the baseline values of the groups involved are equal, or nearly so, a quick review of the data in the literature suggested that this was not the case in the studies of interest. The use of a gain score (which measures the difference between before and after EQ-5D scores for each group) allowed for an unbiased comparison of the groups and was therefore used. The MD included in the equation above is therefore actually the mean difference of differences, as shown below.

$$g = (1 - \frac{3}{4n - 9})\frac{(\mu_1^{post} - \mu_1^{pre}) - (\mu_2^{post} - \mu_2^{pre})}{s_p^2}$$

where μ_1^{post} is the average post-surgery EQ-5D score of the group receiving treatment 1.

Objectives

To answer the research questions identified above, several objectives need to be set. The overall objective of this study is to use NMA analyses to identify the optimal treatment for femoral neck fractures according to EQ-5D, mortality and AVN. This can be further broken down into the following three objectives: i) identify all published studies that meet the selection criteria, ii) conduct NMA and test assumptions, and iii) rank treatments.

Hypotheses

As this study aims to compare treatment effects ($\theta_{treatment}$) to identify the most optimal with regards to three outcomes, three sets of hypotheses are needed.

As previously mentioned, the treatment effect for EQ-5D is the gain score SMD. The null hypothesis for this outcome is therefore as follows: the change in quality of life before and after surgery is equal for each treatment. In numerical form, the null hypothesis is as follows: H_0: $gs_1 = gs_2 = \ldots = gs_n$ where gs_t is the gain score ($\mu_t^{post} - \mu_t^{pre}$) for treatment *t,* from treatments 1 to *n*. The alternative hypothesis on the other hand is as follows: the change in quality of life from pre- to post-surgery is not equal among surgical treatments for displaced femoral neck fractures. In numerical form, it is: H_1: $gs_1 \neq gs_2 \neq \ldots \neq gs_n$. These hypotheses can also be stated in terms of Hedge's g. The null and alternative hypotheses would be: H_0: g = 0 and H_1: g ≠ 1 for every pairwise contrast.

As previously mentioned, the TE for mortality and AVN is the odds ratio (OR), which is the odds that a person affected by the outcome in question received a treatment divided by the odds that someone experienced the outcome but did not receive the treatment. An OR>1 means the treatment is associated with higher odds of outcome, and a OR<1 means the opposite.

The hypotheses for these two outcomes are structured in the same way. The null hypotheses are as follows: the odds of experiencing either death or AVN is equal among all therapies. In numerical form, the null hypothesis are as follows: H_0: $\theta_1 = \theta_2 = \ldots = \theta_n$, where θ_n is the odds of experiencing a fracture in treatment n. The alternative hypotheses on the other hand are as follows: the odds of experiencing one of the outcomes is not equal among treatments. In numerical form, it is: H1: $\theta_1 \neq \theta_2 \neq \ldots \neq \theta_n$. These hypotheses can also be stated in terms of OR. The null and alternative hypotheses would be: H0: OR=1 and H1: OR≠1 for every pairwise contrast, for both outcomes.

It is worth noting that the NMA requires that several assumptions be considered, and each of these can be tested using a statistical analysis.

Methodology

As determined by the objectives discussed above, this NMA is composed of three steps: a literature search, statistical analysis, and ranking of treatments. Although NMAs are a relatively new statistical method (first described in 1997), several guidelines have been published to help conduct and evaluate studies utilizing this statistical method[62]. Those from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and Cochrane were followed during the preparation of this study's methodology[63, 64]. PRISMA was developed in 2009 by an international group of experts in systematic reviews and meta-analyses as an update to its predecessor QUOROM (Quality of Reporting of Meta-Analyses). It offers a checklist of items that should be included in such a study in addition to a standardized flowchart to report the results of the literature review. Both were used in this study. Cochrane on the other hand, offers an entire handbook on how to conduct Cochrane Systematic Reviews and Meta-Analyses, with specific steps to take while conducting the study.

Review protocol registration

As with RCTs, it is a best practice to register and publish protocols for systematic reviews and/or meta-analyses as it can reduce bias, improve the methodology through peer-review and reduce risk of duplication. No review protocol was registered for this study due to time constraints and lack of resources.

Database search

To identify the papers available in the literature, the list of databases searched, and search queries used to find papers of interest within these databases were determined. Once the entire strategy was developed, the Liaison Librarian at the Schulich Library of Physicial Sciences, Life Sciences, and Engineering of McGill University (Nazi Torabi) was consulted to ensure the strategy was optimal. The following databases were included in the literature search:

1. PubMed (1950-current)
2. EMBASE (1947-current)
3. Cochrane Library (1898-current)
4. Clinicaltrials.gov
5. European Union Clinical Trials Register
6. World Health Organization Clinical Trials Registry Platform
7. Web of Science Core Collection (1900-current)
8. Google Scholar

The PICO strategy was used as a backbone for the search queries. However, no outcomes were included in the query in order to identify the largest amount of papers. There was therefore only specification for the population (patients with femoral neck fractures), intervention (fracture treatments), and comparisons (specific treatments to compare). Synonyms were found and combined using the Boolean operator "OR", then combined with other components with the use of the "AND" operator. A combination of free words and indexing terms were used in this process.

All databases were searched using at least free words. Cochrane Library, PubMed and EMBASE also used indexing terms. Search queries for the Cochrane Library and MEDLINE used the MeSH Medical Terms, whereas EMBASE used EMTREE for its indexing terms. Search queries therefore had the following basic structure: "femoral neck fracture" AND "treatment" AND "clinical trial" NOT exclusion terms. All synonyms of "femoral neck fracture" were found and included as free words. Index terms that either alone or combined refer to "femoral neck fracture" were also included and joined to the free words with the use of the "OR" operator. All synonyms of "treatment" and the names and synonyms of specific treatments were included as free words. Index terms for these were also included and joined to the free words with the use of the "OR" operator. To filter for clinical trials, a combination of index terms and free words previously developed for this purpose was referred to. These were found in the Cochrane Handbook for Systematic Reviews of Interventions and have been shown to have adequate sensitivity and specificity[65]. For some databases (ex. EMBASE), there were different combinations available that either optimized specificity, sensitivity, or both. When the choice was available, sensitivity was chosen to minimize the likelihood of missing papers of interest. Finally, some exclusion terms were used. These are listed in Appendix A along with the rest of the search queries, with the number of papers found after each search entry.

In addition to RCTs, systematic reviews and meta-analyses that focussed on the research questions were also identified. The bibliographies of these studies were used as an additional source. Each paper that seemed to fit the topic of interest within these bibliographies were searched for and added to the list of papers of interest.

Selection criteria

Once the databases and systematic reviews/meta-analyses were searched, the title of the results were scanned by one individual (JL) and papers of interest were imported into a citation manager (EndNote X8, Clarivate Analytics). Various import filters were used. These can be found on the Endnote website (links are in Appendix B). Most duplicates were identified and removed using the software's "find duplicates" function. However, many duplicates remained. While some were easy to identify, several were different enough that in depth investigation was required. It was therefore necessary to compare other aspects of the papers.

The results of some studies were published in more than one paper, however with different titles, years and/or first authors. Some were found to have been published twice, using the same results, but with different perspectives (for example, one paper looked at the quality of life, and the other reported the quality of life in addition to cost of treatments to conduct a cost utility analysis). Others, which were harder to identify, published results on participants at a certain point in time, then published results for the same participants, but at a later point in time. Identification of these papers was facilitated by the fact that some authors stated in the introduction or methods section of the paper published at a later date that it was a continuation of a previous study. Some were even harder to identify as they simply added new participants to the study, without stating that it was a continuation of a previous study. These were identified by comparing a multitude of factors. For example, two studies with the same sample size, comparing the same treatments, between groups with the same average age and proportion of women in the samples, would be deemed to be the same study. This unfortunately leads to the need to assume they are the same paper, which is unfavorable; papers should really only be excluded with certainty. However, avoiding inclusion of duplicates was favored in this study and the decision was made to exclude when various factors suggested they were duplicates.

The titles and abstracts of the remaining papers were scanned by the same individual and only papers that seemed to answer the research questions were kept. The inclusion and exclusion criteria were then implemented. The inclusion criteria were as follows:

1. Studies with participants with femoral neck fractures
2. Studies comparing femoral neck fracture treatments
3. Studies with the following outcomes: quality of life, mortality or avascular necrosis
4. Randomized controlled trial
5. Published in peer-review journal accessible to authors

The exclusion criteria were the following:

1. Duplicates (obvious and hidden duplicates)
2. Not in French or English
3. Non-randomized trials
4. Conference presentations

5. Ongoing trials
6. Papers reporting outcomes of interest in a unit that cannot be converted to the desired unit.

The papers included also need to have treatments categorized and reported appropriately. For example, studies comparing bipolar to unipolar hemiarthroplasty but do not specify if cement was used or leave it to the discretion of the surgeon whether to use cement would need to be excluded as it does not fit the model of this study. Furthermore, the papers needed to report the data needed for the quantitative analysis. For example, a paper simply stating a significant difference between two treatments for a certain outcome could not be included, the studies needed to report actual data.

Bias Assessment

The potential for bias in each study used in the NMA was evaluated using the Cochrane Bias Assessment tool. This tool evaluates the following seven areas that can introduce bias in an RCT: random sequence generation, allocation concealment, selective reporting, incomplete outcome data, blinding of participants and personnel, blinding of outcome assessment, and other bias. A description of these sources of bias can be found in Appendix C. This part of the study was conducted by two individuals (JL and JG).

While the Cochrane guideline is helpful, it was necessary to specify further what would deem a paper to have a high, low or unclear risk of bias. While the randomization procedure was easy to categorize, other areas were not as straight forward. As explained in the guideline, selective reporting of outcomes was determined by searching for study protocols. If the protocol did not mention the outcome of interest, the risk of bias was found to be low. If it did mention the outcome, but the final paper reports the outcome using a different unit of measurement, the bias was found to be high, yet if it does report the outcome and with the same unit as expected, the risk was low.

To determine if there are other biases, the papers were individually searched in various databases to determine if letters to the editor or comments had been sent to the editors to notify them of fraudulent claims or other negative feedback. Only negative feedback relating to the data used in this study deemed the paper to be biased.

When the paper was a duplication of previously published work, all titles were searched. Papers were also reviewed to determine if there was a conflict of interest disclosure. A difference of more than 5 patients lost to follow-up between treatment arms merited a high-risk assessment in terms of incomplete outcome data. For the risk of bias assessment for the blinding of personnel to the allocated treatment, if there were more than 6 surgeons involved, then the knowledge of the allocation was not deemed to be a risk of bias. In cases where the EQ-5D was self-assessed, the risk of bias due to lack of blinding of outcome assessment was deemed to be low. Complete blinding of the patient to their treatment allocation, and self-assessment of the EQ-5D resulted in lowest risk.

Data extraction and cleaning

Once the final papers for inclusion were chosen, a spreadsheet was created using Excel (Microsoft, United States) to import the data of interest into. Data for the following variables were extracted for all three outcomes: authors, year of publication, title, inclusion criteria, exclusion criteria, time to follow-up, country where the study was conducted, whether the study used an intention to treat principle, which treatments were compared, number of participants at randomization, number of participants at follow-up, age (average with standard deviation), percentage of participants that are female, mental health score (with name of test), displacement of fracture (displaced, non-displaced or both included). While the majority of this data was not used in a quantitative analysis, it served to give an overview of the studies characteristics; to determine if there are any substantial differences between the studies that could be a source of heterogeneity.

Data extracted only for QoL analysis includes the estimated pre-fracture QoL index scores (with standard deviation) and the post-surgery QoL scores (with standard deviation). Data extracted only for the mortality analysis includes the number of deaths and the number of participants at follow-up. Data extracted solely for the AVN analysis includes the number of AVN cases and the number of patients at follow-up. The data as stratified by the year of follow-up as well as fracture displacement. As such, if the study had multiple follow-ups, data was extracted for each endpoint. Furthermore, if the study included displaced and non-displaced fractures but published data for the two combined and both individually, all possible combinations were extracted into the spreadsheet.

Data for QoL were converted to EQ-5D index scores using algorithms found in the Health Economics Research Centre database[66]. Ultimately, data from only one study (reporting SF-12 data) was converted as others did not report sufficient data. The algorithm used can be found in Appendix D. Only papers reporting QoL indexes with sufficient data to be converted to EQ-5D were included.

One study reporting AVN presented an odds ratio comparing the two treatments, in contrast to all other papers, which reported number of cases. This study was therefore excluded in this first step of the NMA, but then reintroduced in the later step, as explained in greater detail below.

Missing data was imputed in several ways. As there were many duplicate publications of the same study, if the paper that was used to extract outcome data lacked other important data (for example, number of patients at follow-up or standard deviations), its duplicates were reviewed to determine if they reported this missing data. For example, studies that did not estimate the pre-fracture EQ-5D scores were found to have duplicates that did report this data. Missing data on standard deviations of EQ-5D scores at one-year follow-up, which occurred in one study, and missing data on number of participants at follow up were imputed using regression techniques. There was no missing data for mortality and AVN analyses.

Network Meta-Analysis

Direct treatment effect calculations

The next step is to conduct pairwise meta-analyses that calculate the summary direct treatment effects for comparisons covered in more than one study (the term direct refers to the head-to-head comparisons made in an actual study). This treatment effect is calculated by finding the weighed average of the effects observed in each study comparing the same two studies. The weights are assigned according to either sample size or the inverse of the variance (precision) to give more importance to studies with higher power.

Software and model selection

Two statistical models can be used according to the level of variance between treatment effects found in different studies (known as heterogeneity, which will be explained in more detail below); while a fixed effect model assumes all variance is due to random error and is appropriate when there is no or very little heterogeneity, a random effect model assumes variance is due to another source of error (such as conceptual differences between studies) and is appropriate when there is heterogeneity, but to a certain extent. Analyses in this study were conducted using both fixed and random effects models.

Next, a statistical framework was chosen. Two statistical paradigms exist that differ in their philosophy and methodology with regards to their perspective on statistics and probability. They are known as frequentist and Bayesian approaches. While there is an ongoing debate regarding which is best, there is substantial evidence showing that they give similar results in an NMA setting. Fundamentally, the main difference is that the frequentists believe statistical models are fixed and that the data collected from a population changes, and Bayesians believe the data is fixed and that models vary. The most important difference from a methodological standpoint is that a Bayesian analysis allows for prior knowledge of a system to be considered in the analysis (called the “prior”), while a frequentist would have no prior knowledge considered. Analyses were carried out using both approaches for EQ-5D and compared afterwards. Analyses for mortality and AVN were only conducted using frequentist methods. For NMAs where there is no prior, a non-informative prior can be assigned, and since there are no appropriate priors for the analyses conducted in this study, a non-informative prior was used for the Bayesian analysis.

R, an open-source program that allows for both frequentist and Bayesian NMAs was used to calculate indirect treatment effects given the direct treatment effects. The netmeta package was used for the frequentist analyses and the nmaINLA package was used for the Bayesian analysis[67, 68].

Indirect and mixed treatment effect calculations

To summarize this method, consider the treatment effect calculated earlier, μ^{D}_{AB}. If another treatment was compared to A, say C, there will also be a treatment effect μ^{D}_{AC}. If there is no study comparing B and C, an indirect estimation of the treatment effect (μ^{I}_{BC}) can be calculated according to the following equation: $\mu^{I}_{BC} = \mu^{D}_{AC} - \mu^{D}_{AB}$. For treatments that have both indirect and direct treatment effects, a mixed treatment effect (μ^{M}) is measured by taking a weighed average of the two. In summary, NMA gives summary treatment effects for all pairwise comparisons, even if they have not been directly measured in a RCT.

Assumption testing

Several assumptions need to be considered to ensure the validity of results generated in an NMA. Conceptual heterogeneity, which refers to the conceptual differences between trials (ex. study design, participant demographics, and definition of terms used, among others) can lead to differences in the direct and indirect treatment effect estimates, which is known as inconsistency. Furthermore, this can result in statistical heterogeneity, which refers to instances in which different RCTs evaluating the same two treatments report contrasting effect sizes.

Various methods exist to detect whether inconsistency exists in a network. For conceptual heterogeneity, each trial was evaluated by comparing the data collected on each study and differences will be noted; for inconsistency, a node-splitting approach, which compares the indirect and direct estimates, can be used. A significant difference in the two estimates was used to detect inconsistency. Statistical heterogeneity was measured using the Cochran's Q and I^2 statistics, respectively. While the former is useful to detect significant heterogeneity, the latter represents the proportion of variation in treatment effect estimates that is the result of heterogeneity[69]. If the p-score for the Q-statistics is below 0.05, then there is significant heterogeneity. The threshold for I^2 to demonstrate significance is a bit more ambiguous as different authors use varying cut-offs. The threshold used in these analyses was 30%. Once the fit of assumptions has been assessed, treatment were ranked according to their P-scores, a method proposed and validated by Rücker and Schwarzer[70].

Sensitivity analyses

As previously mentioned, NMAs were conducted using using both fixed and random effects models. Analyses for EQ-5D were done using both Bayesian and frequentist methods. In addition to varying the statistical models used, data was also stratified according to several other variables.

There was a total of 6 analyses done for the mortality NMA. Data was stratified by length of follow-up (all, 1 year, and 2-5 years) and treatments were grouped in various ways;

analyses were done with different internal fixation techniques categorized individually (dynamic hip screw, nail, pins, cannulated screws), and other analyses with internal fixation techniques grouped together (grouped as internal fixation. EQ-5D and AVN data was not stratified.

EQ-5D/Mortality Partial Order Set Analysis

The netposet function of the netmeta R package was used to combine the rankings for EQ-5D and Mortality to give a single multi-outcome ranking. This is accomplished using partial order set analysis method. This analysis compares treatment rankings for both outcomes and develops a new order by observing which treatments are better or worst for all outcomes. The logic and exact methodology is explained elsewhere[68, 71, 72]. The results of the partial order set (poset) analysis were summarized using a partial order plot and Hasse Diagram. The plot was generated using the plotting function with the netmeta package and the Hasse diagram was generated by attaching the hasseDiagram package to the netmeta package[73].

Results

The results of this study are composed of three sections for each outcome, and one for the EQ-5D/mortality partial order set analysis. The first is the results of the literature search and is composed of the PRISMA flowchart (combined for all outcomes), a table of the final papers included and their characteristics, some descriptive charts and a network chart. The second is the results of the bias assessment, which is composed of the Cochrane bias assessment tool results. Third is the results of the NMA, which is composed of summary pairwise treatment effects for each comparison, forest plots with the worst treatment as reference (instead of placebo, as it is common practice), ranking of treatments according to p-scores, and heterogeneity and inconsistency test results. The last section is composed of the results of the partial order set for EQ-5D and mortality, which includes the partial order in numerical and graphical form.

Prisma Flow Chart (combined)

Figure 3.1 shows the flowchart for paper selection, from database searching to selection criteria implementation and final paper selection. A total of eight databases were searched (including PubMed, Embase and Cochrane Library) and 10,541 results were found using the search queries listed in Appendix A. 71 systematic reviews were related to this study's research questions. From these reviews, 3470 papers were added to the list. There was therefore a total of 14,011 papers retrieved. After an initial removal of duplicates, conducted using the "find duplicates" function of EndNote X8, 7119 papers remained. Following review of titles and abstracts, and a first implementation of the selection criteria (simply to ensure the papers focussed on the correct research question), 227 papers remained. After a second selection criteria implementation resulted in the removal of another 98 papers (rationales are listed in Figure 3.1 below), 129 papers were left. Of these, 19 reported QoL indexes that were either EQ-5D or convertible to EQ-5D, 94 reported mortality data, and 45 reported AVN data. Of the 19 papers reporting QoL, 10 were either hidden duplicates or did not report sufficient data, leaving only 9 for inclusion into the NMA. Among the 94 that reported mortality data, 66 were either hidden duplicates, had insufficient data, or compared treatment groups that were inappropriate for this study (compared two types of unipolar hip arthroplasties, or two types of internal fixation treatments). Of the 45 papers that reported AVN data, 31 were either hidden duplicates, lacked sufficient data, or grouped patients into treatment groups that was not suitable for this study (compared a treatment to one that cannot lead to AVN, any of the arthroplasties for example).

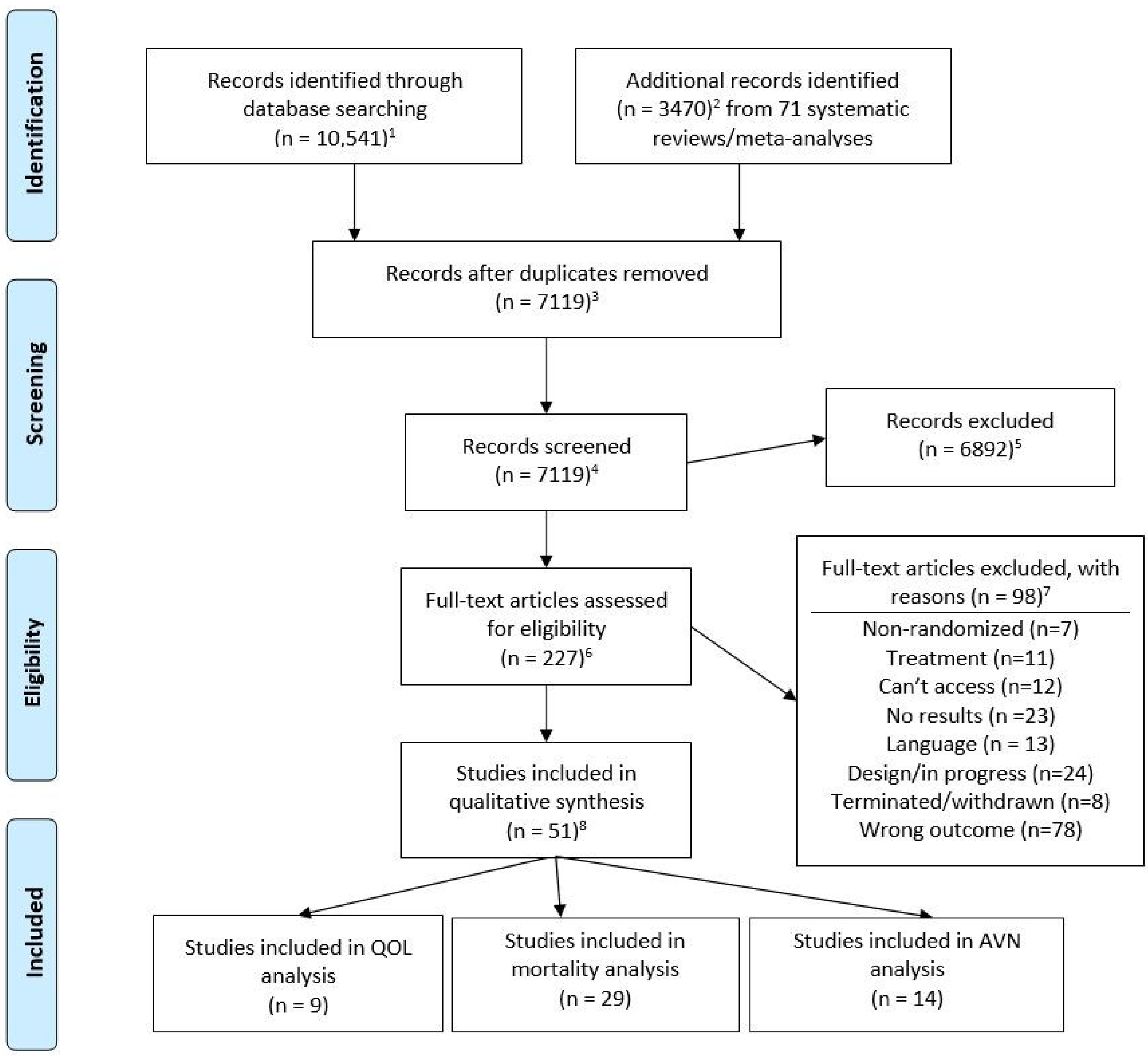


Figure 3.1: Prisma flow chart for systematic review of EQ-5D, mortality and AVN after hip fracture surgery[1]. Databases searched include PubMed (MEDLINE), Embase, Cochrane Library, Web of Science, Scopus, ClinicalTrials.gov, EU Clinical Trials Register, WHO International

Clinical Trials Registry Platform, and Google Scholar. Search queries are shown in Appendix A. [2]Systematic reviews/meta-analyses found in databases. [3]Duplicates found using the EndNote X8 "find duplicates" function. [4]Papers that had their titles and abstracts reviewed. [5]Papers excluded based on their titles and abstracts.[6]Papers fully reviewed. [7]Papers excluded based on full review. [8]Papers that fit inclusion criteria. Final papers chosen excludes hidden duplicates (individual studies with more than one publication), papers with insufficient data (even for imputation).

EQ-5D results

Sample characteristics

There was a total of nine studies that fit the inclusion criteria for the QoL analysis. One was initially included but was ultimately removed as it led to significant inconsistency. This paper only had data at two years and included only patients with dementia, which may explained why it introduced such inconsistency. Eight compared two treatments, and one compared three. Each study was composed of a unique combination of treatments being compared. The network of comparisons is shown in Figure 3.2, with the red triangle showing the three-arm trial. There was a considerable number of hidden duplicates; the study labelled Blomfeldt was found to have two[74, 75] Hedbeck_2 was to have one[76], Langslet was found to have one[77], Frihagen was found to have two[78, 79], and Inngul was found to have one[80].

Data on potential covariates was extracted and reviewed to ensure there was no substantial conceptual heterogeneity. The earliest publication was from 2005 and the most recent was from 2017. All studies were conducted in northern Europe; the majority took place in Sweden (see Figure 3.4). Only one study (Langslet) did not use an intention to treat principle. Average ages and sex distribution were comparable. As shown in Figure 3.3, the treatment with the highest number of participants was BPHAC (517 patients), whereas THAU only had 34 participants. As demonstrated in Figure 1.2, the comparison with the highest and lowest precision were IF:BPHAC and IF:BPHAU, respectively. Although this study did not purposefully only include studies that included patients with displaced femoral neck fractures, all studies found excluded undisplaced femoral neck fractures. These findings are important to note as they will impact the clinical interpretation of the results. Of the studies that included IF as a treatment, three used cannulated screws (Blomfeldt, Chammout, and Hedbeck_2) and one gave the surgeon the choice between cannulated screws and dynamic hip screws.

Table 3.1: Characteristics of studies used in EQ-5D; first author, year of publication, country of publication, treatments compared, sample size at randomization (N_{ran}), average age and standard deviation (NR; not reported) and percentage of each group that is female (%fem) of included studies are listed. Moerman study reported SF-12 QOL values, which were converted to EQ-5D. All studies have 1yr follow-ups for EQ-5D. Acronyms for treatments are as follows: THAC; total hip arthroplasty cemented, IF; internal fixation, THAU; total hip arthroplasty uncemented, BPHAC; bipolar hemiarthroplasty cemented, BPHAU; bipolar hemiarthroplasty uncemented, UPHAC; unipolar hemiarthroplasty cemented.

Study label	Year	Country	Treatments	N_{ran}	Average age (SD)	%fem
Blomfeldt	2005	Sweden	THAC/IF	49/53	79(5)/81(7)	82/79
Chammout	2017	Sweden	THAC/THAU	35/34	72(4)/73(5)	63/74
Frihagen	2007	Norway	BPHAC/IF	110/112	83(7)/83(7)	71/78
Hedbeck_1	2011	Sweden	BPHAC/THAC	60/60	81(5)/80(5)	90/78
Hedbeck_2	2013	Sweden	UPHAC/IF	29/30	85(5)/84(5)	83/83
Inngul	2013	Sweden	BPHAC/UPHAC	60/60	85(NR)/87(NR)	70/82
Keating	2006	Scotland	BPHAC/THAC/IF	65/66/65	75(7)/75(6)/74(7)	78/75/74
Langslet	2014	Norway	BPHAC/BPHAU	112/108	83(6)/83(6)	78/74
Moerman	2017	Netherlands	UPHAC/UPHAU	110/91	83(6)/84(6)	75/67

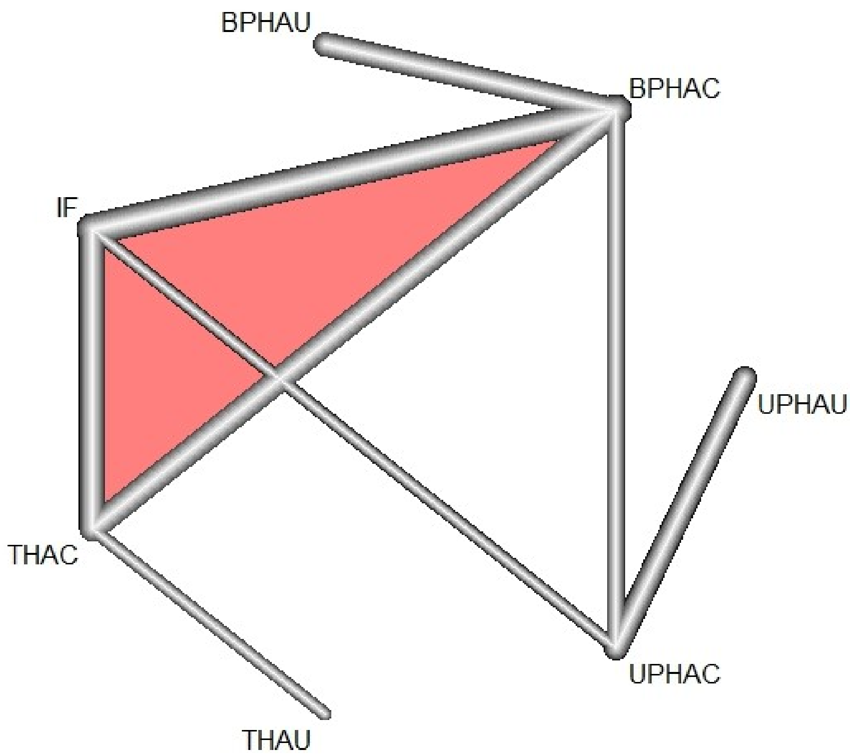


Figure 3.2: Network graph for EQ-5D RCTs. Edges represent comparisons made in the literature; nodes represent treatments. Red triangle indicates the three-arm study. Thickness of line indicates precision (thicker line indicates higher precision). Acronyms for treatments are as follows: THAC; total hip arthroplasty cemented, IF; internal fixation, THAU; total hip arthroplasty uncemented, BPHAC; bipolar hemiarthroplasty cemented, BPHAU; bipolar hemiarthroplasty uncemented, UPHAC; unipolar hemiarthroplasty cemented.

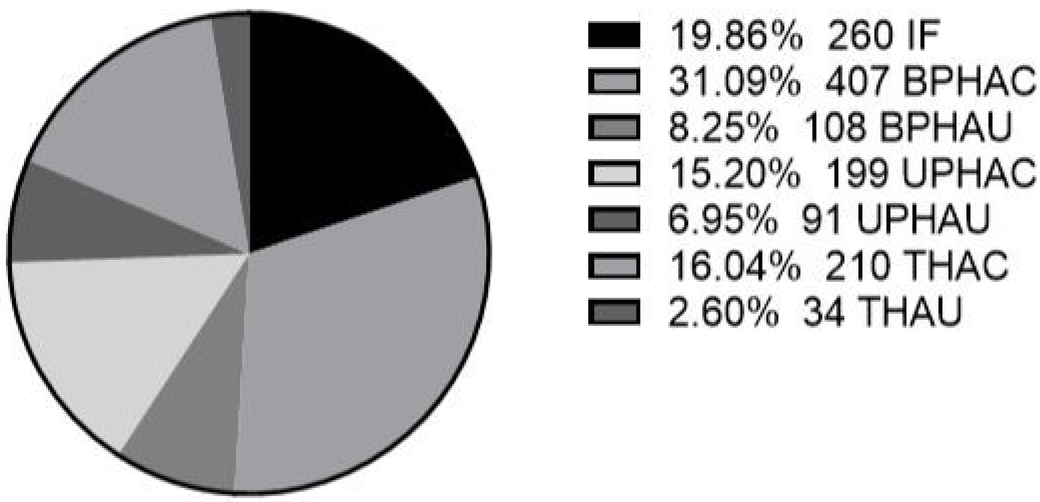


Figure 3.3: Overall sample sizes of each treatment at randomization. Acronyms for treatments are as follows: THAC; total hip arthroplasty cemented, IF; internal fixation, THAU; total hip arthroplasty uncemented, BPHAC; bipolar hemiarthroplasty cemented, BPHAU; bipolar hemiarthroplasty uncemented, UPHAC; unipolar hemiarthroplasty cemented. The order of treatments in the legend (top to bottom) corresponds to the colors in the pie chart in a clockwise order.

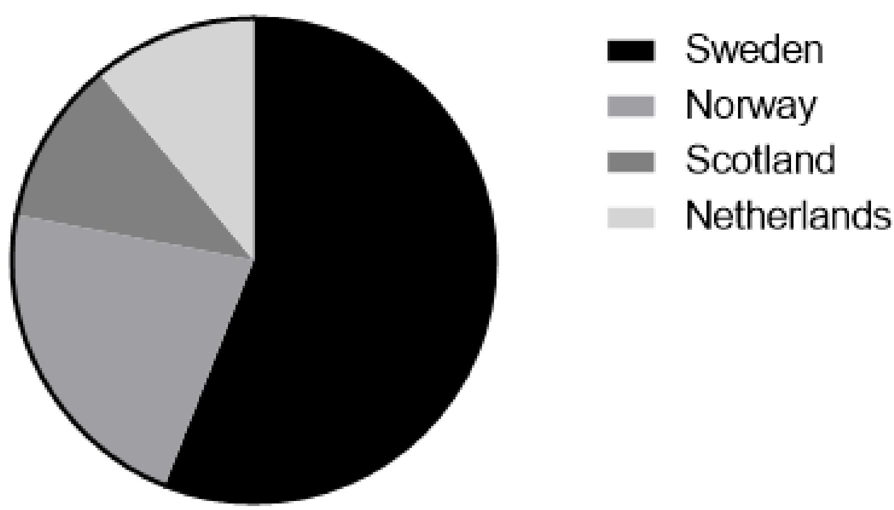


Figure 3.4: Countries authoring the studies included in the EQ-5D NMA.

Bias Assessment

Results of the bias assessment show that the risk of bias was mediocre. Only 54% of the assessed characteristics were deemed to have low risk of bias. All papers were appropriately randomized but due to the nature of the studies, blinding of the surgeons and patients was impossible and therefore all studies were scored as having a high risk of bias for the blinding of the participants and personnel. The surgeon is without a doubt aware of the surgery being performed, and the patient is most likely to see their x-rays. Nonetheless, as this category requires blinding of both, the fact that the surgeon was aware merits this category a high risk of bias assessment. 78% of the studies blinded the outcome assessor. The one paper that was found to have a “other bias” stated in their paper that four of the authors had conflicts of interest.

Table 3.2: Risk bias assessment of studies included in EQ-5D analysis using the Cochrane Risk Bias Assessment tool. See Appendix C for guidelines on how papers were evaluated. Two individuals completed the assessment individually. Results were compared and differences were discussed until ratings were unanimously agreed upon.

Study label	**Random sequence Generation`**	**Allocation concealment**	**Blinding of participants and personnel**	**Blinding of outcome assessment**	**Incomplete outcome data**	**Selective reporting**	**Other bias**
Blomfeldt	1	1	0	0	1	?	1
Chammout	1	1	0	1	1	1	1
Frihagen	1	1	0	1	?	1	1
Hedbeck1	1	1	0	0	?	?	?
Hedbeck2	1	1	0	0	0	1	1
Inngul	1	1	0	0	1	?	?
Keating	1	1	0	0	?	1	1
Langslet	1	1	0	0	1	?	0
Moerman	1	1	0	0	0	1	?

NMA, Assumption Tests and Ranking

Results of the NMA (shown in Table 3.3 and Figure 3.5) found 8 significant contrasts. These are highlighted in bold in Table 3.3. IF was found to be the worst treatment and was therefore set as the reference (instead of a placebo). BPHAC, THAC and THAU were found to be significantly better than the reference. The large confidence interval for the IF-THAU comparison should be noted. Within, between and overall heterogeneity were found to be insignificant (p=0.875). Higgins I^2 was therefore set to zero and a fixed effect analysis was reported here. As the I^2 is set to zero, the results of the random effects analysis were identical to those of the fixed effect analysis. The netheat plot shown in Figure 3.6 shows that there was very little inconsistency in the network. As shown in Table 3.5, THAU was ranked highest, with a P-score of 0.975, and IF was ranked lowest, with a P-score of 0.087.

trasts for EQ-5D (gain score SMD h 95%CI) from estimated pre-fracture EQ-5D
. Values in bold are significant. Ac ynms for treatments are as follows: THAC;
IF; internal fixation, THAU; total h arthroplasty uncemented, BPHAC; bipolar
AU; bipolar hemiarthroplasty unc ented, UPHAC; unipolar hemiarthroplasty

	BPHAC	BPHAU	IF	THAC	THAU	UPHAC	UPHAU
BPHAC		**0.283 [0.018 to 0.549]**	**0.335 [0.148 to 0.522]**	-0.201 [-0.418 to 0.017]	**-0.556 [-1.079 to -0.033]**	0.045 [-0.255 to 0.345]	0.068 [-0.341 to 0.476]
BPHAU	**-0.283 [-0.549 to -0.018]**		0.052 [-0.273 to 0.377]	**-0.484 [-0.827 to -0.141]**	**-0.84 [-1.426 to -0.253]**	-0.239 [-0.639 to 0.162]	-0.216 [-0.703 to 0.272]
IF	**-0.335 [-0.522 to -0.148]**	-0.052 [-0.377 to 0.273]		**-0.536 [-0.758 to -0.313]**	**-0.891 [-1.417 to -0.366]**	-0.29 [-0.61 to 0.029]	-0.267 [-0.691 to 0.156]
THAC	0.201 [-0.017 to 0.418]	**0.484 [0.141 to 0.827]**	**0.536 [0.313 to 0.758]**		-0.356 [-0.832 to 0.12]	0.245 [-0.11 to 0.601]	0.268 [-0.183 to 0.719]
THAU	**0.556 [0.033 to 1.079]**	**0.84 [0.253 to 1.426]**	**0.891 [0.366 to 1.417]**	0.356 [-0.12 to 0.832]		**0.601 [0.007 to 1.195]**	0.624 [-0.032 to 1.28]
UPHAC	-0.045 [-0.345 to 0.255]	0.239 [-0.162 to 0.639]	0.29 [-0.029 to 0.61]	-0.245 [-0.601 to 0.11]	**-0.601 [-1.195 to -0.007]**		0.023 [-0.255 to 0.301]
UPHAU	-0.068 [-0.476 to 0.341]	0.216 [-0.272 to 0.703]	0.267 [-0.156 to 0.691]	-0.268 [-0.719 to 0.183]	-0.624 [-1.28 to 0.032]	-0.023 [-0.301 to 0.255]	

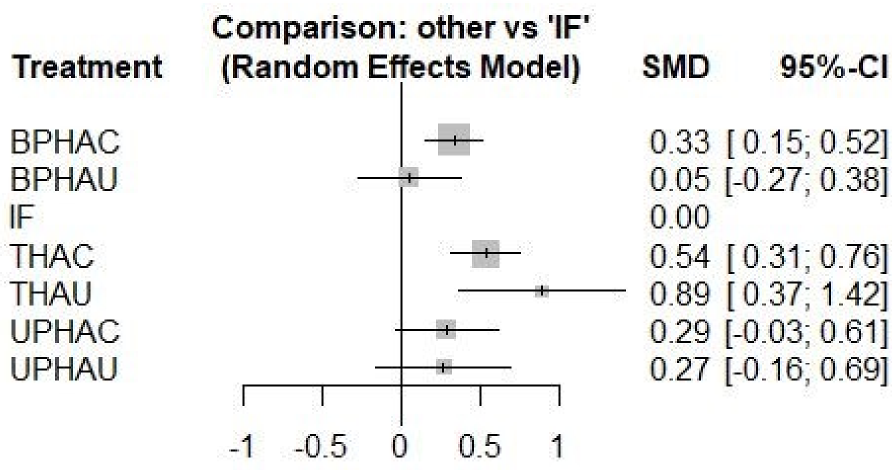


Figure 3.5: Forest plot for EQ-5D network with IF as reference. The size of the grey squares is proportional to the number of participants in that treatment group across all studies. Acronyms for treatments are as follows: THAC; total hip arthroplasty cemented, IF; internal fixation, THAU; total hip arthroplasty uncemented, BPHAC; bipolar hemiarthroplasty cemented, BPHAU; bipolar hemiarthroplasty uncemented, UPHAC; unipolar hemiarthroplasty cemented.

Table 3.4: Results of tests of overall heterogeneity (Q_{total}), within designs (Q_{within}), and inconsistency between designs ($Q_{between}$), with the degrees of freedom (d.f.) and significance (p-value) for each analysis. Higgins I^2 was set to zero by the netmeta package as heterogeneity was not found to be significant (d.f. and p-value are therefore non-applicable – NA).

Q	Value	d.f.	P-value
Q_{total}	1.22	4	0.8755
Q_{within}	0	0	N/A
$Q_{between}$	1.22	4	0.8755
I^2	0		N/A

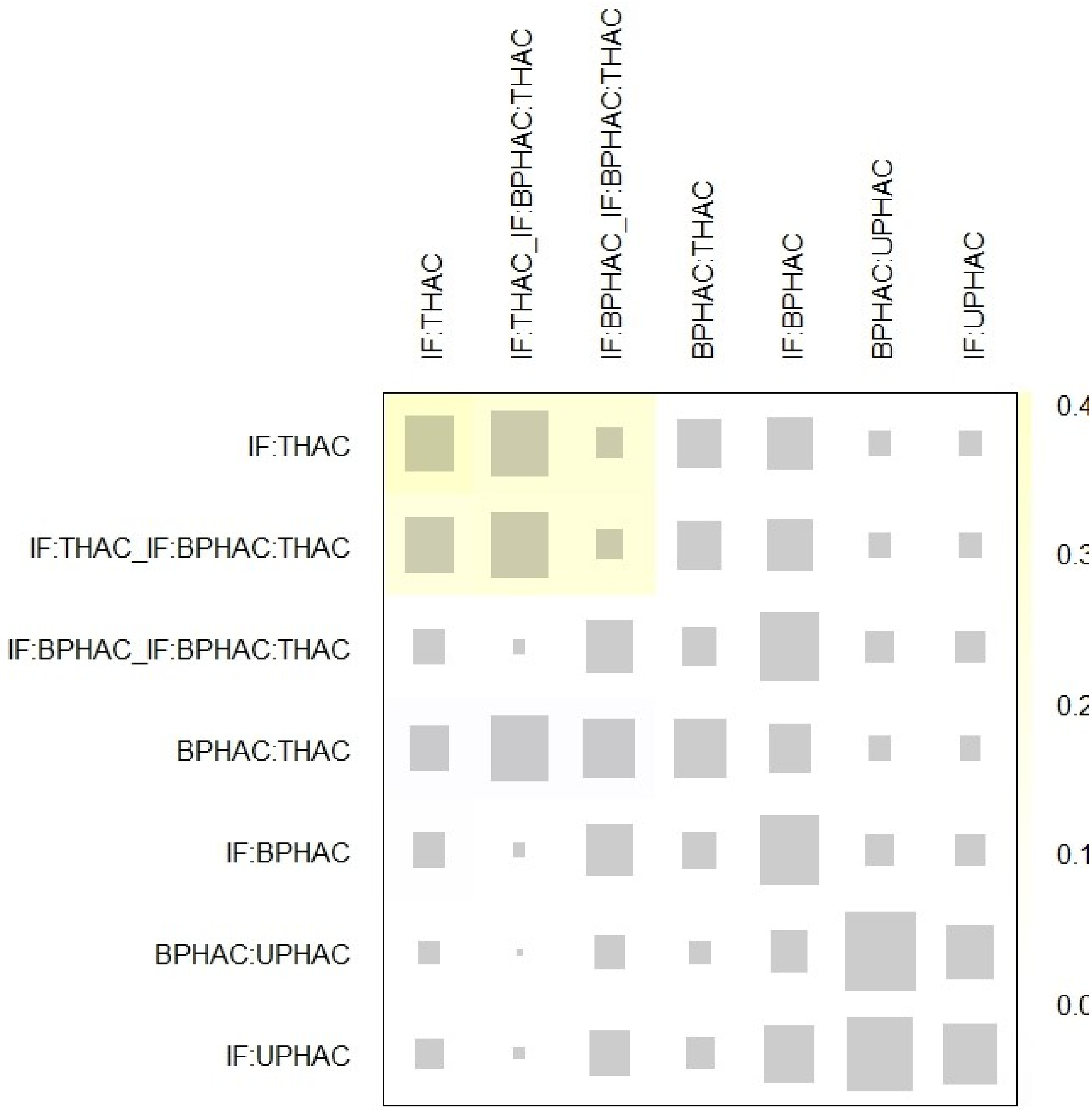


Figure 3.6: Netheat plot for EQ-5D network. Demonstrates the contribution of each design to each estimate in the network and the extent of inconsistency. The size of squares is proportional to the contribution from the comparison in the column to the treatment comparison in the row. Color demonstrates the level of inconsistency in a row comparison when the column comparison is removed from the network. Blue demonstrates when inconsistency increases when the column is detached, and yellow/red demonstrates when inconsistency decreases when the column is detached.

Table 3.5: Ranking of treatments according to one-year EQ-5D gain scores. Greater P-score is indicative of a better treatment.

Treatment	P-score
THAU	0.975
THAC	0.804
BPHAC	0.546
UPHAC	0.484
UPHAU	0.443
BPHAU	0.160
IF	0.087

Mortality

Since the results of the EQ-5D analysis only included papers that had a follow-up period of one year and excluded patients with undisplaced FNF, a decision was made to alter the exclusion criteria for the mortality analysis. While an analysis could be conducted that would include studies with longer follow-up times and/or undisplaced fractures if they did not have a significant level of statistical heterogeneity or inconsistency, the eventual objective of combining the results of the QoL and mortality NMA into an overall ranking suggested that this would not be appropriate. While there are statistical analyses to ensure studies have low heterogeneity among them within a single outcome analysis, the impossibility of such a method to compare between outcomes suggested that a more prudent approach be taken. Studies that included displaced fractured and those with a follow-up time above 1yr were therefore excluded.

Sample Characteristics

A total of 29 studies were identified for inclusion in the mortality analysis. There were two three-arm trials, which are shown in the network chart (Figure 3.9) with blue and red triangles. There was a total of 33 pairwise comparisons, with 11 unique combinations of treatments. As with EQ-5D, there was a considerable number of hidden duplicates. Dynamic hip screws, cannulated screws, nails, targon nails, and pins were combined into the IF group. Seven studies could not be used in this analysis as it compared internal fixation techniques. One study was excluded as it reported contradictory mortality statistics throughout the paper. Eight studies only reported mortality overall and did not report it for each treatment. Five studies were excluded as they compared treatments that were too specific for our research question (for example, short vs. long screws, cannulated screws with or without calcium phosphate augmentation, etc.).

Like the studies included in the EQ-5D analysis, the majority of studies originated from northern Europe; 79% of studies were conducted in either Sweden, Norway, Netherlands, or United Kingdom. The earliest publication was from 1991 and the most recent was from 2017; only three publications occurred prior to the year 2000 however. Eight studies had 30 patients or less randomized to each treatment arm. The group with the highest number of participants among all studies included was IF, and that with the lowest number is THAU (34 participants). The study labelled Blomfeldt had one[75], Calder had two[81, 82], Figved had one[83], Hedbeck1 had one[76], Inngul had one[80], Parker2 had one[84], Ravikumar had one[85] and van den Bekerom had two[86, 87].

Analysis of potential covariates showed that the studies included were comparable. Average age of patients ranges from 67.3 to 87.4 years. The average age of patients was between 75 and 85 years in 81% of studies that did report age (2 did not report). Sex distributions were similar; 81% of studies that reported sex distribution (2 did not report) had groups composed of 70%-90% women. There were some outliers however; proportions that were female ranged from 47.6% to 97.0%. A large difference between groups within a single trial was observed in the study Ravikumar (18.1% difference). Just over half (52%) of the included studies reported that they used an intention-to-treat principle.

ample size at randomization is given by N_{ran}. NR stands for not reported. Data
forward slash. When data was only available for both groups combined, but not
c is presented.

Authors	Year	Country	Treatment	Nran	Average age (SD)	Sex (% female)
Bachrach	2000	Sweden	THAC/IF	50/50	84.2(5.7)/84(5.3)	80.0/76.0
Blomfeldt	2005	Sweden	UPHAU/IF	30/30	84(5.9)/83.6(5.9)	93.0 (overall)
Calder	1996	UK	BPHAC/UPHAC	118/132	NR	85.6/86.4
Chammout1	2012	Sweden	THAC/IF	43/57	78(NR)/79(NR)	88.0/72.0
Chammout2	2017	Sweden	THAC/THAU	35/34	72(4)/73(5)	62.9/73.5
Davison	2001	England	BPHAC/UPHAC/IF	97/90/93	NR	74.2/78.9/75.3
Deangelis	2012	US	UPHAC/UPHAU	66/64	81.8(9.0)/82.8(7.6)	78.8/75.0
El-Abed	2005	UK	UPHAU/IF	NR	74(NR)/72(NR)	64.5/70.0
Emery	1991	UK	BPHAC/BPHAU	27/26	78(7.2)/79.6(8.0)	88.9/84.6
Figved	2009	Norway	BPHAC/BPHAU	115/115	83.4(5.7)/83(6.3)	78.0/74.0
Frihagen	2007	Norway	BPHAC/IF	110/112	82.5(7.3)/83.2(7.7)	71.0/78.0
Hedbeck1	2011	Sweden	BPHAC/THAC	60/60	80.5(5.1)/80.5(5.1)	90.0/78.0
Hedbeck2	2013	Sweden	UPHAC/IF	30/30	85.2(5.5)/83.8(5.5)	83.0/83.0
Inngul	2013	Sweden	BPHAC/UPHAC	60/60	85.5(NR)/87.4(NR)	70.0/82.0
Moerman	2017	Netherlands	UPHAC/UPHAU	110/91	83(6.2)/84(6.7)	75.0/67.0
Neander	1997	Sweden	IF/THAC	10/10	86 (overall)	75 (overall)
Parker1	2015	UK	UPHAC/IF	26/30	81.2(NR)/81.5(NR)	NR
Parker2	2010	UK	UPHAU/IF	229/226	82.4(NR)/82.2(NR)	80.0/80.0
Parker3	2010	UK	UPHAC/UPHAU	200/200	83(NR)/83(NR)	80.0/73.0
Parker4	2000	Sweden	UPHAU/IF	106/102	82(NR)/81(NR)	NR
Puolakka	2001	Finland	UPHAC/IF	15/17	82(NR)/81(NR)	93.3/76.5
Raia	2003	US	BPHAC/UPHAC	115 (total)	82.4(NR)/81.8(NR)	76.4/58.3
Ravikumar	2000	UK	THAC/UPHAC/IF	89/91/91	81.0(NR)/82.0(NR)/79.7(NR)	90.0 (overall)
Santini	2005	Italy	BPHAC/BPHAU	106 (total)	82.09(7.6)/79.7(8.6)	75.5/79.2
Somashekar	2013	India	BPHAC/UPHAC	21/20	67.3(NR)/75.6(NR)	85.0/47.6
Taylor	2012	New Zealand	UPHAC/UPHAU	80/80	85.3(7.0)/85.1(6.6)	66.3/71.3
van den Bekerom	2010	Netherlands	BPHAC/THAC	137/115	80.3(NR)/82.1(NR)	84/78
van Dortmont	2000	Netherlands	UPHAC/IF	29/31	84(NR)/84(NR)	76.0/97.0
Vugt	1993	Netherlands	BPHAC/IF	22/21	76.0(3.0)/75.3(3.0)	63.6/52.4

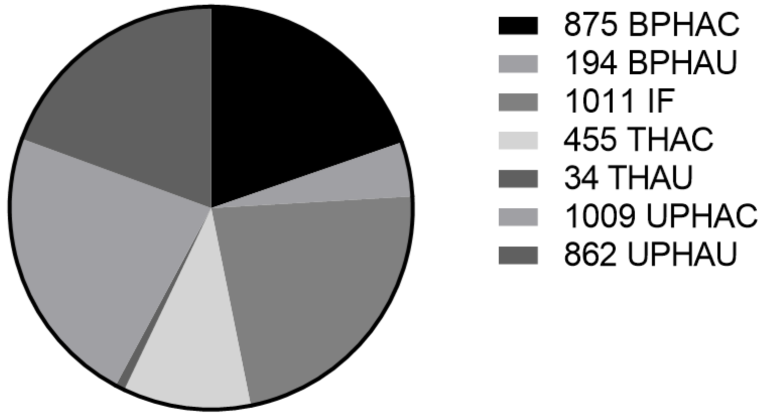


Figure 3.7: Overall sample sizes of each treatment at randomization. Acronyms for treatments are as follows: THAC; total hip arthroplasty cemented, IF; internal fixation, THAU; total hip arthroplasty uncemented, BPHAC; bipolar hemiarthroplasty cemented, BPHAU; bipolar hemiarthroplasty uncemented, UPHAC; unipolar hemiarthroplasty cemented. The order of treatments in the legend (top to bottom) corresponds to the colors in the pie chart in a clockwise order.

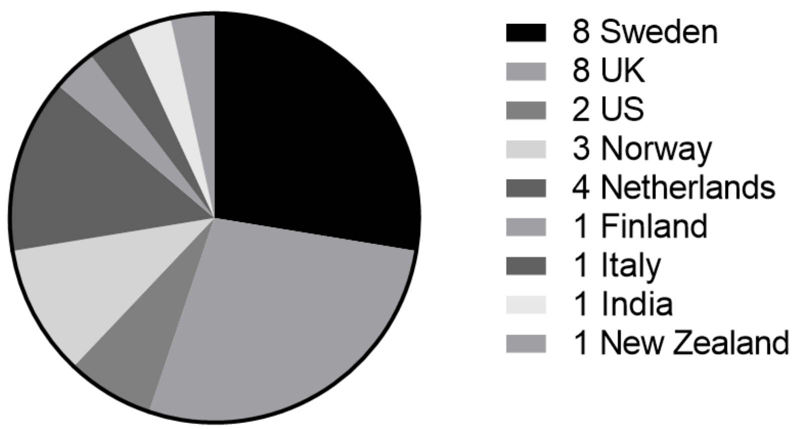


Figure 3.8: Countries authoring the studies included in the EQ-5D NMA. The order of countries in the legend (top to bottom) corresponds to the colors in the pie chart in a clockwise order.

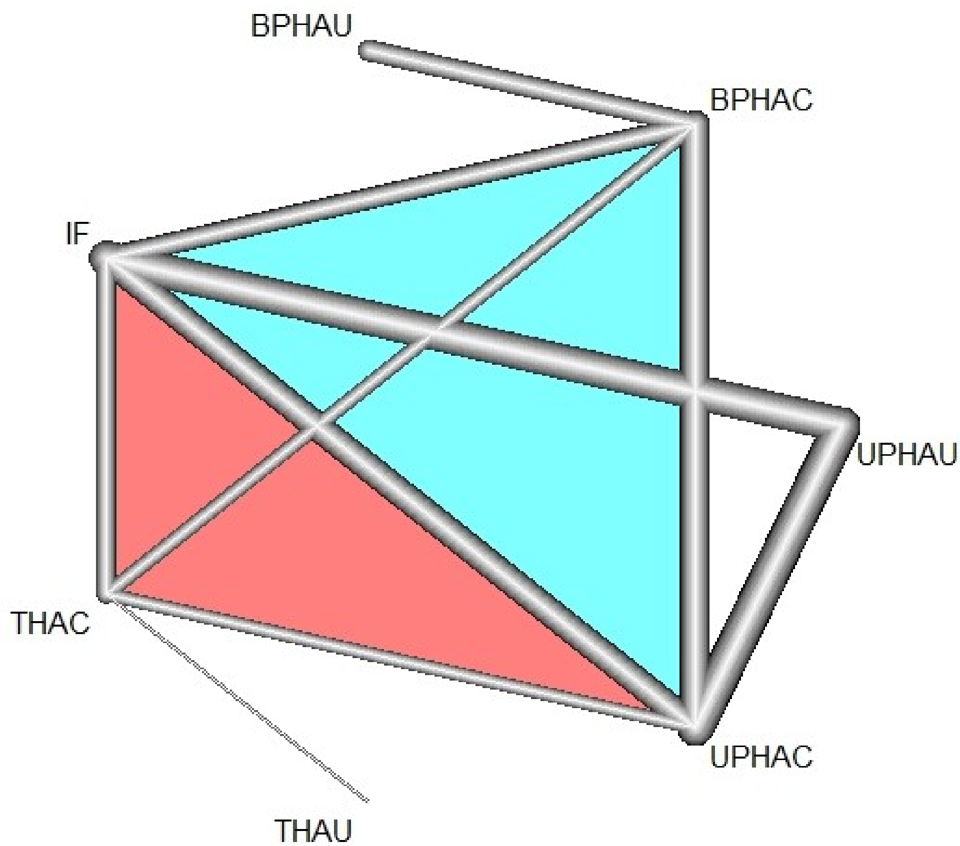


Figure 3.9: Network graph for mortality analysis. Edges represent comparisons made in the literature; nodes represent treatments. Blue and red triangles indicates three-arm study. Thickness of line represents precision (inverse of variance) of the measures treatment effect.

Bias Assessment

The papers included in the mortality analysis were found to have a mediocre level of risk of bias. Low risk of bias was found for 43% of the categories analyzed. Nearly all studies used the appropriate randomization process however many did not report their protocol and were therefore judged to have an unknown risk of bias. Due to the nature of this study, all studies were found to have a high risk of bias with regards to the blinding of personnel. The two studies reported to have a risk of bias for “other bias” stated that the authors had a conflict of interest. Two studies (Santini and El-Abed) were found to have a high risk of bias in most categories.

Table 3.7: Risk bias assessment of studies included in mortality analysis using the Cochrane Risk Bias Assessment tool. See Appendix C for guidelines on how papers were evaluated. Two individuals completed the assessment individually. Results were compared afterward, and differences were discussed until both reviewers agreed on the ratings.

Study label	**Random sequence Generation`**	**Allocation concealment**	**Blinding of participants and personnel**	**Blinding of outcome assessment**	**Incomplete outcome data**	**Selective reporting**	**Other bias**
Bachrach	?	?	0	0	1	?	1
Blomfeldt	1	1	0	0	1	?	1
Calder	?	?	0	?	?	?	1
Chammout1	0	1	0	0	1	1	1
Chammout2	1	1	0	1	1	1	1
Davison	1	1	0	1	?	1	?
Deangelis	1	?	0	0	1	1	1
Emery	1	1	0	?	?	?	?
Figved	1	1	0	0	0	?	1
Frihagen	1	1	0	1	1	1	1

Hedbeck1	1	1	0	0	1	1	?
Hedbeck2	?	1	0	0	0	?	1
Inngul	1	1	0	0	1	1	1
Moerman	1	1	0	0	0	1	?
Parker1	1	1	0	1	1	?	1
Parker2	1	1	0	1	1	?	0
Parker3	1	1	0	0	1	?	1
Parker4	1	1	0	?	1	?	1
Puolakka	1	1	0	?	1	?	1
Raia	1	0	0	0	1	?	?
Ravikumar	?	?	0	0	?	?	?
Santini	0	0	0	0	1	?	1
Somashekar	?	0	0	0	1	?	?
Söreide	0	0	0	0	1	?	1
Taylor	1	1	0	1	0	?	0
Vugt	?	0	0	0	1	?	?
van Dortmont	?	0	0	0	1	?	?
El-Abed	0	0	0	0	1	?	?

NMA, Assumption Testing and Ranking

Results of the NMA are shown in Table 3.8 and Figure 3.10. There were no significant findings. BPHAU was set as the reference since it was found to be the worst. BPHAC, THAC and THAU were found to be significantly better than the reference. Again, the large confidence interval for the BPHAU-THAU contrast should be noted. Within, between and overall heterogeneity were found to be insignificant (p=0.903). Higgins I^2 was therefore set to zero and a fixed effect analysis was reported here. As with the EQ-5D analysis, since I^2 is set to zero, the results of the random effects analysis were identical to those of the fixed effect analysis.

Figure 3.11 shows that there is some inconsistency in the IF-UPHAC comparison.

As shown in Table 3.9, UPHAC was ranked highest, with a P-score of 0.744, and BPHAU was ranked lowest, with a P-score of 0.188. If was ranked second with a P-score of 0.709, similar to that of BPHAU.

aring hip fracture treatments according to mortality OR, shown with 95%
treatments are as follows: THAC; total hip arthroplasty cemented, IF; internal
y uncemented, BPHAC; bipolar hemiarthroplasty cemented, BPHAU; bipolar
AC; unipolar hemiarthroplasty cemented.

	BPHAC	BPHAU	IF	THAC	THAU	UPHAC	UPHAU
BPHAC		0.747 [0.465 to 1.199]	1.141 [0.819 to 1.589]	1.111 [0.72 to 1.715]	0.523 [0.044 to 6.285]	1.159 [0.852 to 1.578]	1.018 [0.709 to 1.462]
BPHAU	1.34 [0.834 to 2.151]		1.528 [0.857 to 2.724]	1.488 [0.783 to 2.829]	0.7 [0.056 to 8.804]	1.553 [0.882 to 2.733]	1.364 [0.751 to 2.475]
IF	0.877 [0.629 to 1.221]	0.654 [0.367 to 1.167]		0.974 [0.642 to 1.478]	0.458 [0.038 to 5.493]	1.016 [0.78 to 1.324]	0.892 [0.694 to 1.148]
THAC	0.9 [0.583 to 1.389]	0.672 [0.353 to 1.278]	1.027 [0.677 to 1.558]		0.471 [0.041 to 5.445]	1.043 [0.68 to 1.602]	0.916 [0.581 to 1.445]
THAU	1.913 [0.159 to 22.992]	1.428 [0.114 to 17.95]	2.182 [0.182 to 26.156]	2.125 [0.184 to 24.589]		2.217 [0.185 to 26.629]	1.947 [0.161 to 23.495]
UPHAC	0.863 [0.634 to 1.174]	0.644 [0.366 to 1.133]	0.984 [0.755 to 1.282]	0.958 [0.624 to 1.471]	0.451 [0.038 to 5.417]		0.878 [0.683 to 1.129]
UPHAU	0.982 [0.684 to 1.41]	0.733 [0.404 to 1.331]	1.121 [0.871 to 1.442]	1.091 [0.692 to 1.721]	0.514 [0.043 to 6.198]	1.139 [0.886 to 1.464]	

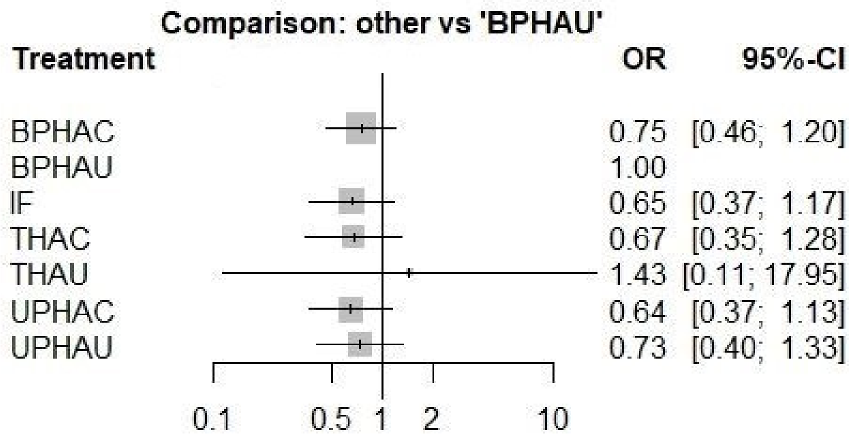


Figure 3.10: Forest plot for mortality network with BPHAU as reference.

Table 3.9: Ranking of treatments according to mortality ORs

Treatment	P-score
UPHAC	0.744
IF	0.709
THAC	0.636
UPHAU	0.464
BPHAC	0.459
THAU	0.301
BPHAU	0.188

Table 3.10: Tests of heterogeneity (within designs) and inconsistency (between designs):

Q measure	Value	d.f.	p-value
Q_{total}	16.38	25	0.903
Q_{within}	13.08	18	0.787
$Q_{between}$	3.3	7	0.856
I^2	0.00	NA	NA

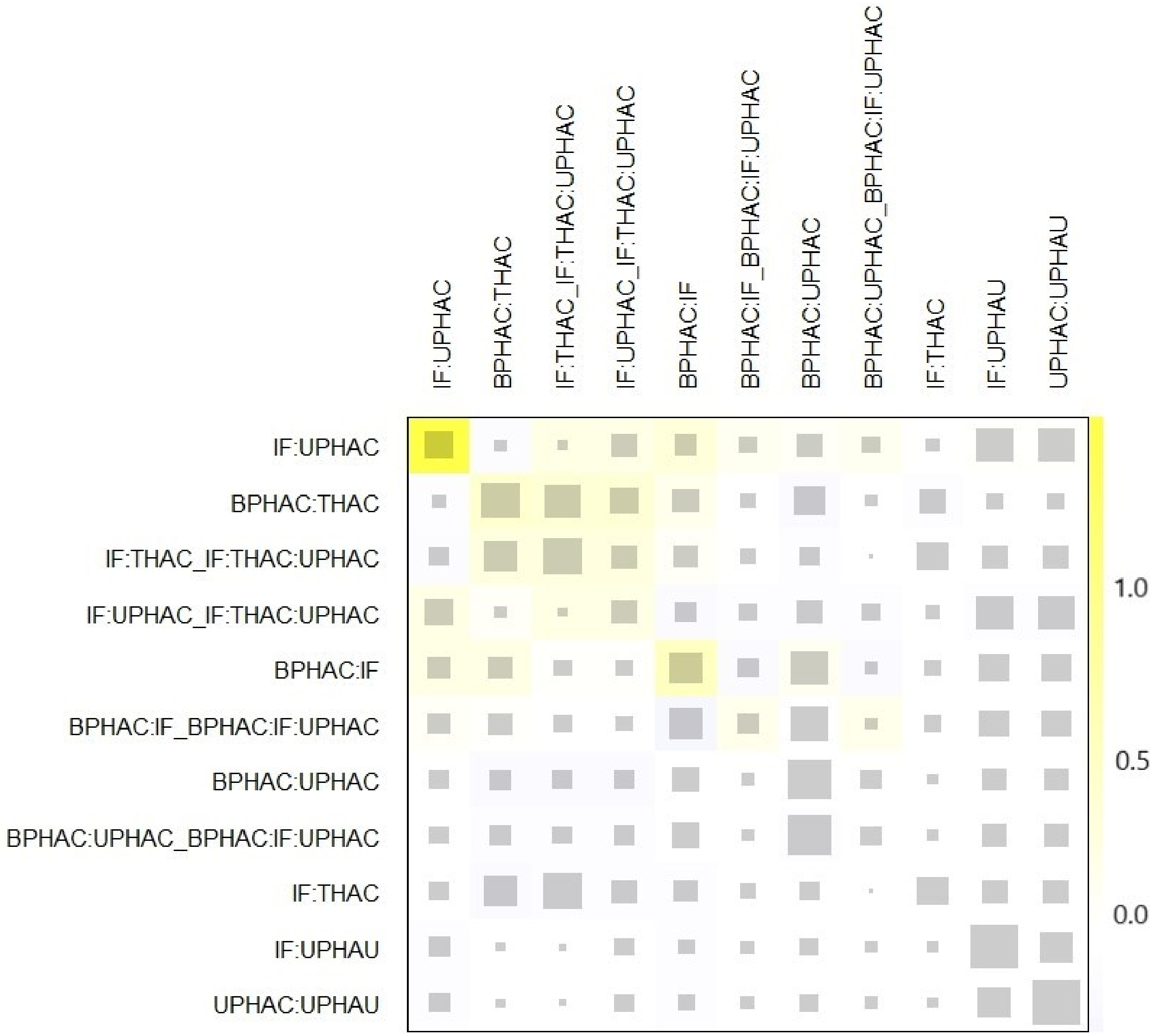


Figure 3.11: Netheat plot for mortality network. Demonstrates the contribution of each design to each estimate in the network and the extent of inconsistency. The size of squares is proportional to the contribution from the comparison in the column to the treatment comparison in the row. Color demonstrates the level of inconsistency in a row comparison when the column comparison is removed from the network. Blue demonstrates when inconsistency increases when the column is detached, and yellow/red demonstrates when inconsistency decreases when the column is detached.

AVN

Sample characteristics

A total of 14 studies were included in the AVN analyses. There were only two hidden duplicates excluded. Frandsen had one[88] and Christie had one[89]. Since arthroplasty treatments cannot lead to AVN, only studies that compared IF treatments were included and treatments were therefore not combined into a single IF group. A total of eight different treatments were found to be compared in the literature. Treatments are listed in the heading of Table 3.11.

The studies included have a wide range of publication dates however. The oldest study was from 1981 and the latest was dated 2017. Six of the studies did not report average ages, and concerning sex distribution, another six either did not report any statistic or only reported a statistic for the entire sample, not per group. Two studies have considerably lower numbers of women compared to the other studies reviewed thus far (Lykke, Siavashi and Liu). CS and Pins had the highest numbers of participants (518 and 505, respectively), and SNP had the lowest (19). SNP, CSC, CSIG and CSFG have small sample sizes that should be noted. The CS/Pins comparison had the largest number of studies comparing them, as well as precision, as shown by the thickness of the edge connecting the two treatments in figure 3.12. Seven studies had data for displaced fractures only, six studies had data for displaced and undisplaced combined, and one study had only undisplaced fractures included. Two studies had data for 1yr follow-up, twelve studies had follow-up periods longer than 1yr, ranging from 15 months to 5 years. It is worth noting that the average age of the samples was substantially lower than those in the previous analyses. Since younger individuals require a more severe trauma to fracture their femoral neck, the trauma that caused these cases was potentially more severe than in other studies and is therefore a potential covariate that could affect outcomes.

Table 3.11: First author, year of publication, country of publication (SCT, Scotland; IRN, Iran; IND, India; CHI, China) , treatments (DHS, Dynamic Hip Screw; CS, Cannulated Screws; SNP, Sliding Nail Plate; CSC, CS with Cement; CSFG, CS with Fibular Graft; CSIG, CS with Iliac Graft), sample size at randomization (N_{random}), average age with SD or range (NR, Not Reported), and % of participants that are female in the studies used in AVN NMA.

First author	Year	Country	Treatment	N_{random}	Average age (SD or Range)	% female
Christie	1988	Scotland	Pins/DHS	66/61	NR	NR
Frandsen	1981	Denmark	Nail/DHS	196/187	NR (28-96)/NR (22-95)	78
Herngren	1992	Sweden	CS/Pins	96/84	77 (32-96)/78 (28-97)	64/62
Holmberg	1990	Sweden	Pins/Nail	110/110	79 (NR)/78 (NR)	75/75
Kuokkanen	1991	Finland	CS/DHS	16/17	72.5 (62-82)/60 (21-84)	NR
Lykke	2003	Norway	CS/Pins	131/147	81 (56-96)/82 (27-101)	24/24
Mjørud	2006	Norway	CS/Pins	101/98	81 (12)/80 (10)	77/76
Nordkild	1985	Denmark	DHS/SNP	90/19	NR	70/63
Siavashi	2017	Iran	CS/DHS	28/30	28 (18-58)/30 (18-60)	25/17
Sørensen	1992	Denmark	CS/DHS	38/35	77(10)/76(9)	71/79
Paus	1986	Norway	CS/DHS	65/66	NR	82/82
Mattsson	2006	Sweden	CS/CSC	60/58	NR (60-98)	83
Kumar	2015	India	CS/CSFGF	45/42	NR (20-50)	NR
Liu	2015	China	CS/CSIG	34/31	34 (9)/35 (9)	39

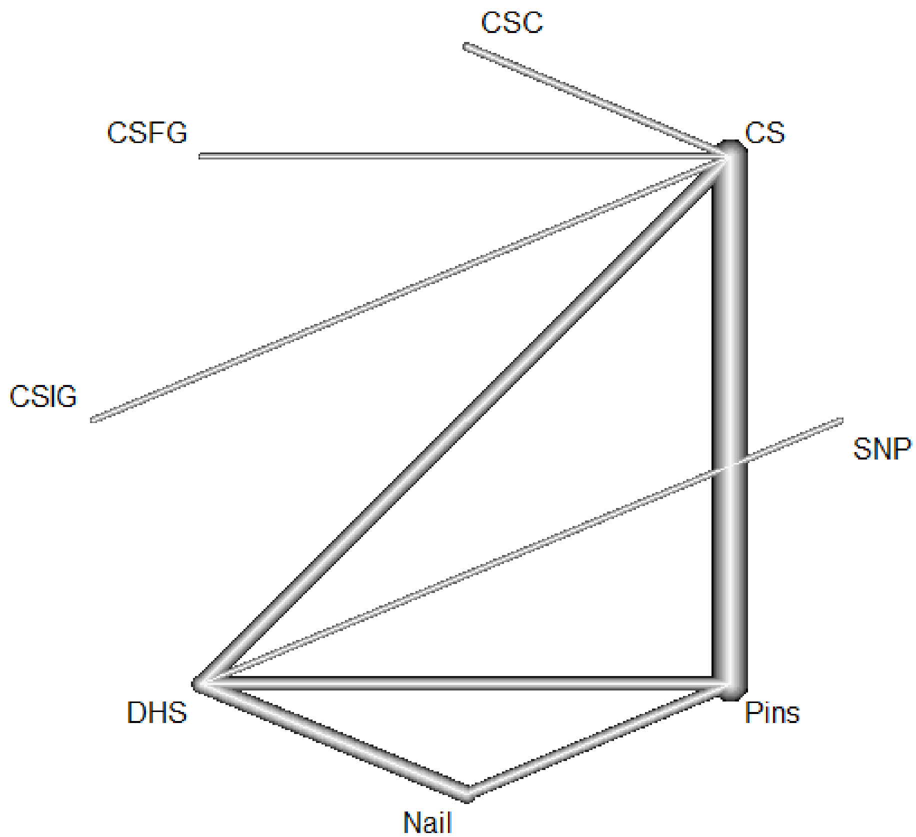


Figure 3.12: Network graph for AVN. Edges represent comparisons made in the literature; nodes represent treatments. Thickness of line indicates precision (inverse of variance).

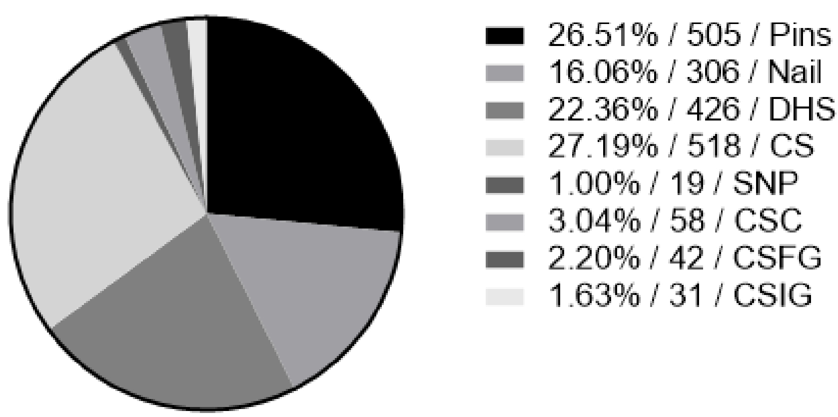


Figure 3.13: Overall sample sizes of each treatment. The order of treatments in the legend (top to bottom) corresponds to the colors in the pie chart in a clockwise order.

Bias Assessment

The risk of bias for the studies included in the AVN was deemed to be high. Just over a quarter (27%) of categories were found to have low risk of bias. 100% of studies were found to have an unknown risk of bias as none of the studies had published protocols for their studies prior to commencement. It was therefore impossible to determine if they purposefully excluded certain outcomes. One study was found to have a high level of risk in the majority of categories (Frandsen).

Table 3.12: Risk bias assessment of studies included in mortality analysis using the Cochrane Risk Bias Assessment tool

Study label	Random sequence Generation`	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
Christie	?	?	0	?	1	?	1
Frandsen	0	0	0	0	?	?	1
Herngren	1	?	0	0	1	?	1
Holmberg	?	?	0	?	?	?	1
Kumar	1	0	0	?	1	?	1
Kuokannen	?	?	0	?	1	?	1
Liu	?	?	0	?	1	?	1

Lykke	1	1	0	0	?	?	1
Mattsson	1	1	0	?	0	?	0
Mjorud	1	1	0	?	1	?	?
Nordkild	?	?	0	?	0	?	1
Paus	?	?	0	?	1	?	1
Siavashi	?	?	0	?	1	?	?
Sorensen	?	?	0	?	0	?	0

NMA, Assumption Testing and Ranking

NMA results for AVN found five significant contrasts, which are shown in Table 3.13. Within, between and overall heterogeneity were found to be insignificant (p=0.467). Higgins I^2 was therefore set to zero and a fixed effect analysis was reported here. As with the previous two outcomes analyzed, since I^2 is set to zero, the results of the random effects and fixed effect analyses were identical. As shown in the netheat plot (Figure 3.15), the DHS:Nail and Nail:Pins comparisons contributed to some minor inconsistency in the Nail:Pins comparison. Furthermore, the CS:DHS and CS:Pins reduced inconsistency in the DHS:Pins comparison and the DHS:Pins comparison reduced inconsistency in the CS:DHS comparison.

CSIG was ranked highest with a P-score of 0.938 and CSC was ranked lowest with a P-score of 0.049. CSC was therefore set as the reference in the forest plot shown in Figure 3.14, which shows that three treatments were found to be significantly better that CSC (CSIG, DHS, and Pins), with ORs of 18.46, 4.7 and 4.36, respectively. SNP and DHS were found to have quite similar P-scores (0.631 and 0.624, respectively).

	CS	CSC	CSFG	CSIG	DHS	Nail	Pins	SNP
CS		0.35 [0.09 to 1.41]	1.43 [0.23 to 9.00]	**6.55 [1.29 to 33.26]**	1.67 [0.95 to 2.93]	1.31 [0.65 to 2.64]	**1.54 [1.07 to 2.23]**	2.03 [0.37 to 11.31]
CSC	2.82 [0.71 to 11.21]		4.03 [0.4 to 40.22]	**18.46 [2.19 to 155.71]**	**4.7 [1.06 to 20.88]**	3.69 [0.78 to 17.36]	**4.36 [1.04 to 18.16]**	5.73 [0.63 to 51.85]
CSFG	0.70 [0.11 to 4.41]	0.25 [0.02 to 2.48]		4.58 [0.39 to 53.4]	1.17 [0.17 to 8.0]	0.92 [0.13 to 6.57]	1.08 [0.17 to 7.06]	1.42 [0.11 to 17.62]
CSIG	**0.15 [0.03 to 0.78]**	**0.05 [0.01 to 0.46]**	0.22 [0.02 to 2.54]		0.25 [0.05 to 1.42]	0.20 [0.03 to 1.17]	0.24 [0.04 to 1.25]	0.31 [0.03 to 3.30]
DHS	0.60 [0.34 to 1.05]	**0.21 [0.05 to 0.94]**	0.86 [0.13 to 5.87]	3.93 [0.70 to 21.93]		0.78 [0.44 to 1.41]	0.93 [0.53 to 1.62]	1.22 [0.24 to 6.17]
Nail	0.76 [0.38 to 1.54]	0.27 [0.06 to 1.27]	1.09 [0.85 to 7.83]	5.00 [0.85 to 29.4]	1.27 [0.71 to 2.30]		1.18 [0.61 to 2.3]	1.55 [0.28 to 8.72]
Pins	**0.65 [0.45 to 0.93]**	**0.23 [0.06 to 0.96]**	0.93 [0.14 to 6.05]	4.24 [0.80 to 22.44]	1.08 [0.62 to 1.89]	0.85 [0.43 to 1.65]		1.32 [0.24 to 7.32]
SNP	0.49 [0.09 to 2.74]	0.17 [0.02 to 1.58]	0.70 [0.06 to 8.71]	3.22 [0.30 to 34.24]	0.82 [0.16 to 4.15]	0.64 [0.11 to 3.61]	0.76 [0.14 to 4.22]	

nparing hip fracture treatments according to AVN OR, shown with 95% confidence
ghted in bold. Acronyms for treatments are as follows: CS; cannulated screws, CSC;
nt, CSFG; cannulated screws with fibular graft, CSIG; cannulated screws with iliac
; sliding nail plate.

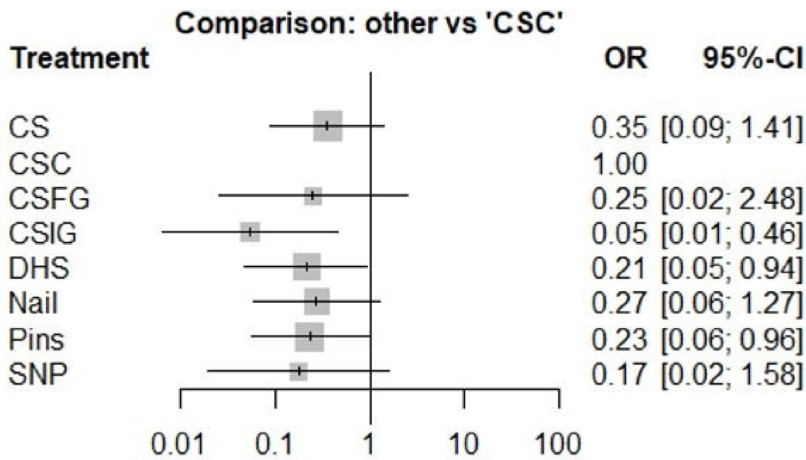


Figure 3.14: Forest plot for EQ-5D network with IF as reference.

Table 3.14: Ranking of treatments according to AVN ORs.

Treatment	P-score
CSIG	0.938
SNP	0.631
DHS	0.624
Pins	0.572
CSFG	0.496
Nail	0.437
CS	0.254
CSC	0.049

Table 3.15: Tests of heterogeneity (within designs) and inconsistency (between designs):

Test	Values	d.f.	P-value
Q_{total}	6.64	7	0.47
Q_{within}	3.63	5	0.6
$Q_{between}$	3.01	2	0.22
I^2	0	NA	NA

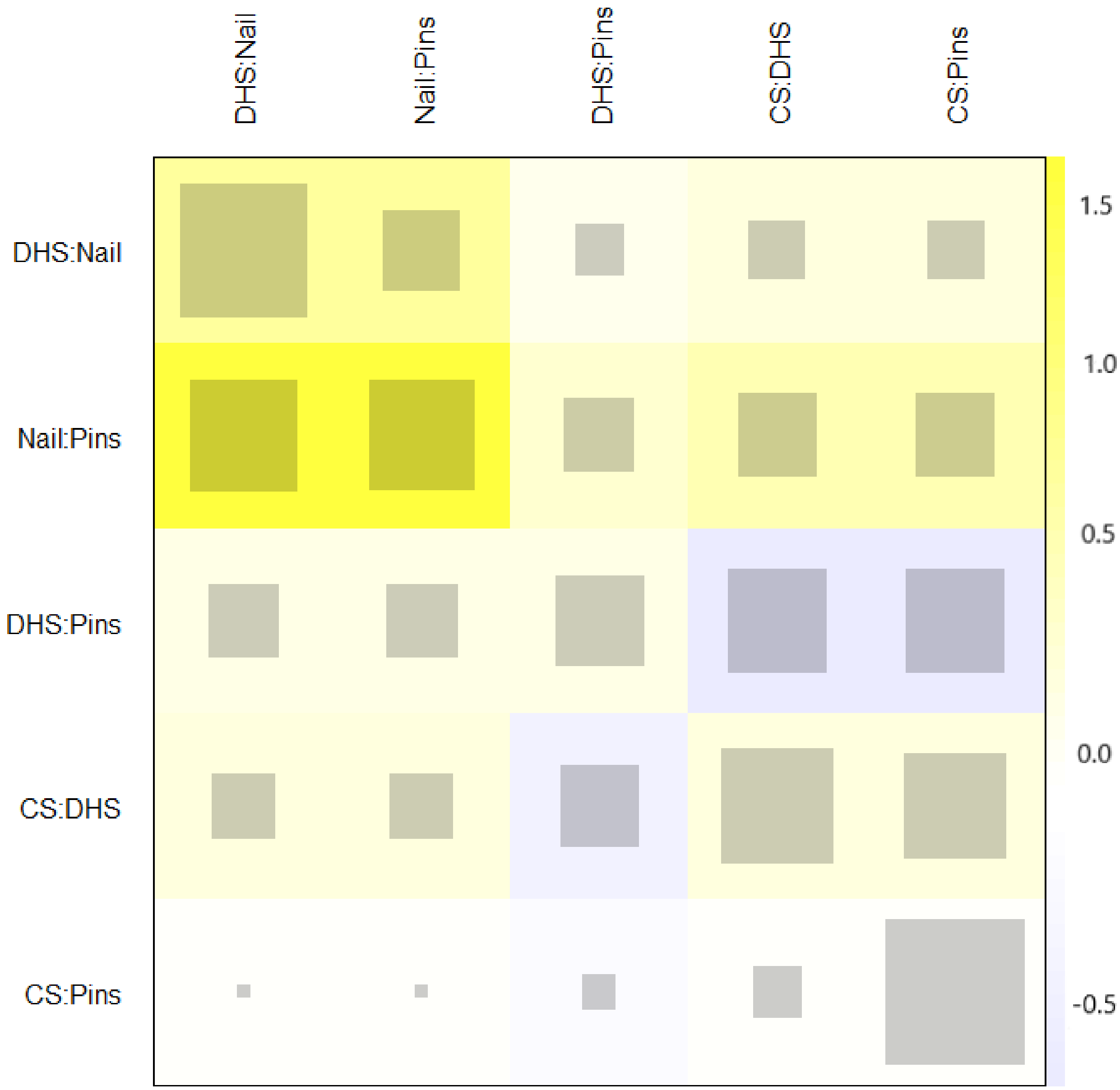


Figure 3.15: Netheat plot for AVN network. Demonstrates the contribution of each design to each estimate in the network and the extent of inconsistency. The size of squares is proportional to the contribution from the comparison in the column to the treatment comparison in the row. Color demonstrates the level of inconsistency in a row comparison when the column comparison is removed from the network. Blue demonstrates when inconsistency increases when the column is detached, and yellow/red demonstrates when inconsistency decreases when the column is detached.

EQ-5D/Mortality Partial Order Set Analysis

The results of the poset analysis are shown below. A table showing the relation between treatments is shown in Table 3.16 From this data, a Hasse diagram was generated and is shown in Figure 3.17. A 2-axis plot showing the rankings of each treatment for each outcome along each axis was also generated and is shown in Figure 3.16.

Table 3.16: Poset results for the NMA comparing bipolar hip arthroplasty with and without cement (BPHAC and BPHAU, respectively), unipolar hip arthroplasty with and without cement (UPHAC and UPHAU, respectively), total hip arthroplasty with and without cement (THAC and THAU), and internal fixation (IF) for EQ-5D and mortality at 1yr follow-up after displaced femoral neck fracture. A "1" indicates that the treatment in the column to the left is immediately better than the treatment directly above it, whereas "0" indicates that it is either worse or incomparable.

	BPHAC	BPHAU	IF	THAC	THAU	UPHAC	UPHAU
BPHAC	0	1	0	0	0	0	0
BPHAU	0	0	0	0	0	0	0
IF	0	0	0	0	0	0	0
THAC	1	0	0	0	0	0	1
THAU	0	1	0	0	0	0	0
UPHAC	0	0	1	0	0	0	1
UPHAU	0	1	0	0	0	0	0

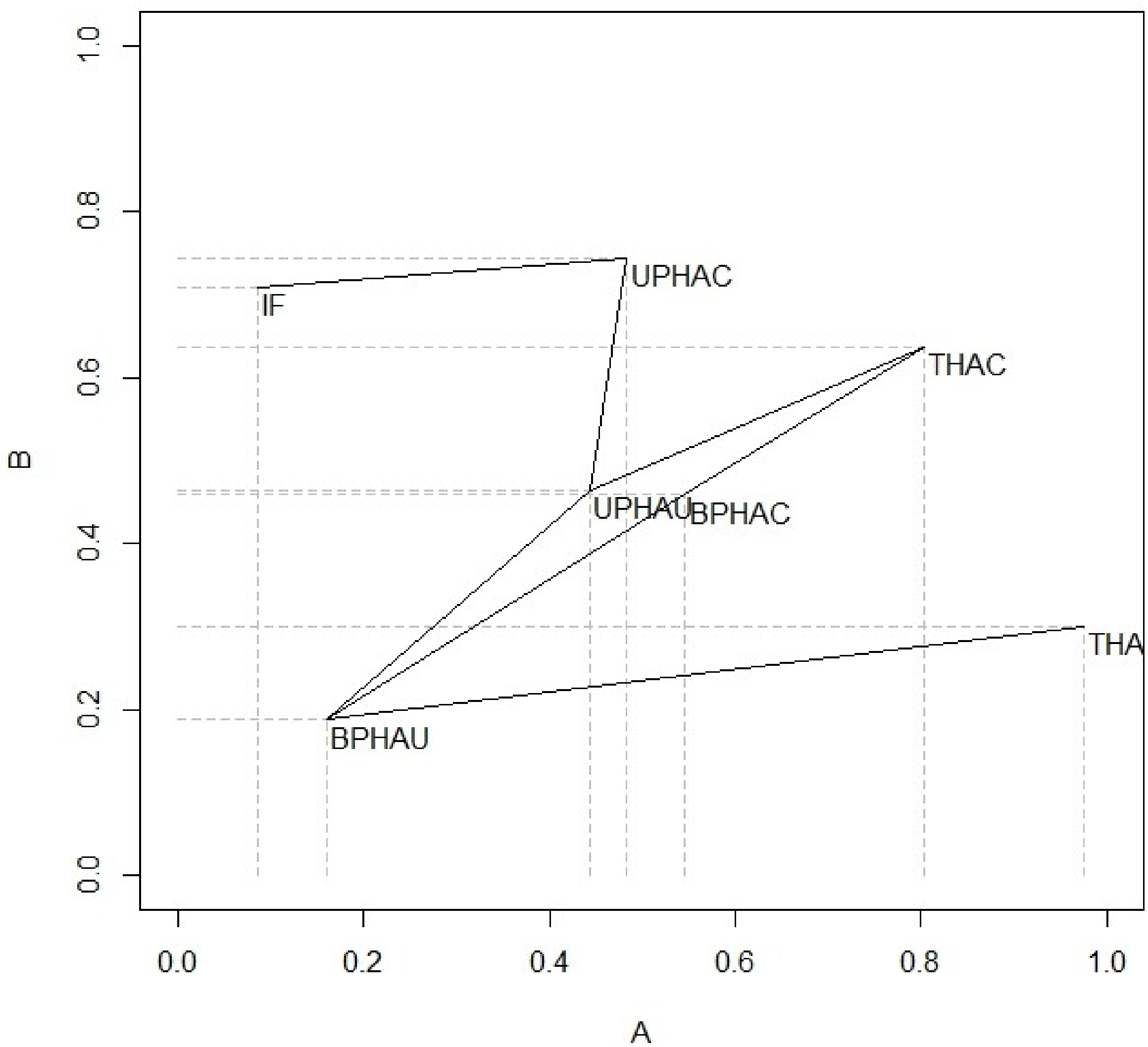


Figure 3.16: Poset plot from the results of the NMA comparing bipolar hip arthroplasty with and without cement (BPHAC and BPHAU, respectively), unipolar hip arthroplasty with and without cement (UPHAC and UPHAU, respectively), total hip arthroplasty with and without cement (THAC and THAU), and internal fixation (IF; either cannulated screws, dynamic hip screw, pins and nail) for EQ-5D and mortality at 1yr follow-up after displaced femoral neck fracture. Ranks were determined by P-scores calculated using a frequentist approach.

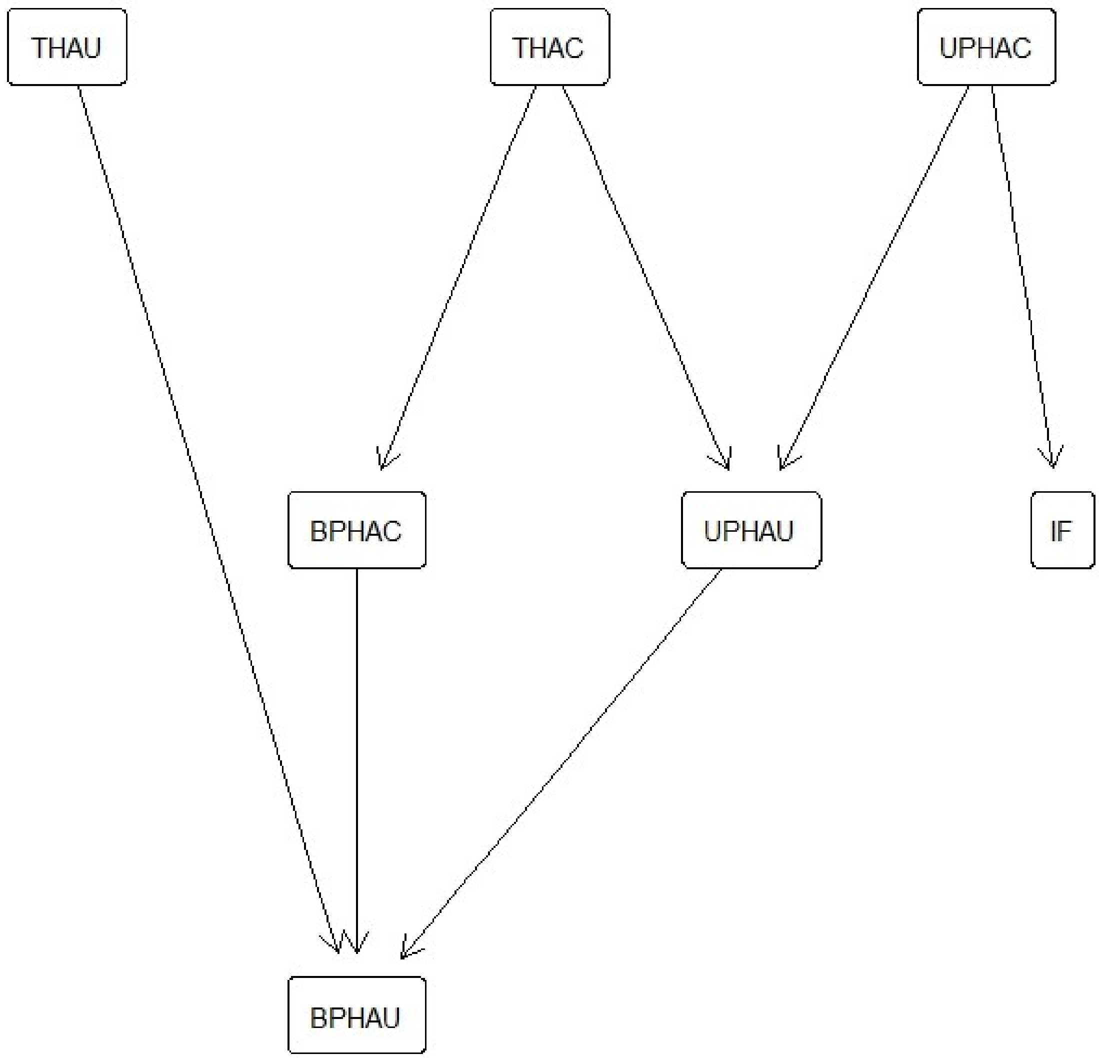


Figure 3.17: Hasse diagram from the results of the NMA comparing bipolar hip arthroplasty with and without cement (BPHAC and BPHAU, respectively), unipolar hip arthroplasty with and without cement (UPHAC and UPHAU, respectively), total hip arthroplasty with and without cement (THAC and THAU), and internal fixation (IF; either cannulated screws, dynamic hip screw, pins and nail) for EQ-5D and mortality at 1yr follow-up after displaced femoral neck fracture. Individual outcome ranks were determined by P-scores calculated using a frequentist approach. Treatments higher in the diagram are better than those below them. Treatments not connected by an arrow are incomparable.

Discussion

The ultimate purpose of this study was to identify the optimal treatments for femoral neck fractures according to EQ-5D, mortality and AVN, then develop a combined ranking for EQ-5D and mortality. As shown in the section above, these objectives have been met and will now be discussed in further detail below; this will include an interpretation of the results, how this study contributes to the literature and how it differs from other similar studies, limitations and solutions, and future studies that will advance knowledge further.

Interpretations and original contributions

To interpret the results of this study, there is first a need to put them into context. Although the selection criteria of this study do not focus on a specific demographic, the clinical and methodological particularities of the papers included should be considered – recall that the results of this study are representative of the population the data originates from. As previously described, the studies included are composed of elderly, mostly women with displaced FNF. In fact, the analysis on EQ-5D only included studies with patients suffering displaced FNF and had one-year follow-ups. Therefore, while the results show that THAU is best and IF is worst, this is only supported by data for displaced fractures at one year for example.

There is also a need to consider the treatment effect unit used, which was the SMD of gain scores for EQ-5D. There are a couple rules of thumb to interpret results reporting the SMD, or more specifically Hedge's g. First, when g = 1, the means of the two groups being compared differ by one standard deviation, when g = 2, they differ by 2 standard deviations, and so on. Second, a major contributor to the development of SMD measures, Jacob Cohen, defines an SMD = 0.2 to constitute a small effect, SMD = 0.5 to be a medium effect and SMD > 0.8 to be a large effect for the social sciences [90]. Cohen nonetheless cautions that stating that an effect size is small or large does not confer any importance about the outcome. For example, a small effect when speaking of mortality is invaluable and should not be dismissed.

The fact that the SMD is standardized implies that it has been adjusted for the variance in scales and precision of the measurements [91]. The SMD therefore controls for the fact that studies included used different forms of the EQ-5D score (3L or 5L) and that they came from countries with variable valuation systems. Regarding scales, as mentioned when outlining research questions, the SMD is a dimensionless treatment effect measure that can incorporate data collected using different scales. Interpretations of the results for the EQ-5D should therefore avoid stating that the results are indicative of a difference in EQ-5D specifically, but rather simply that there is a difference in terms of QoL. In terms of precision, it controls for the fact that the variance in each study may differ, which may be due to the samples' varying characteristics or the studies different methods for evaluating the QoL score [90]. Furthermore, there is a need to recall that the SMD used was for gain scores and that the figure reported are thus differences of differences. An SMD of 1 would indicate that there is a difference of 1 standard deviation in the difference from pre to post-operation between the two treatments.

In addition to context, there is a need to review the importance of significance testing when interpreting the rankings found. Significance was defined as a treatment effect with a 95%CI that does not contain 0. While the P-scores calculated represent the probability that each treatment is best, and not second, third, fourth and so on, the pairwise contrasts and their significance need to also be considered. The table below is taken from the results for EQ-5D above, with brackets added to indicate significant contrasts. Therefore, while THAU is ranked higher than THAC, there is no significant difference between the two and this should inform the interpretation of this ranking.

Table 4.1: Ranking of treatments according to one-year EQ-5D gain scores. Greater P-scores are indicative of a better treatment. Brackets show significant differences according to 95% CI of pairwise contrasts.

Treatment	P-score
THAU	0.9754
THAC	0.8039
BPHAC	0.5462
UPHAC	0.4837
UPHAU	0.4434
BPHAU	0.1601
IF	0.0873

The following are a few important contrasts that will be focussed on in this discussion due to their clinical importance: arthroplasties vs. IF, cemented vs. uncemented implants, bipolar vs. unipolar hemiarthroplasties, hemiarthroplasties vs. total hip arthroplasties. Other interesting findings are nonetheless highlighted. While these are relevant for the QoL and mortality analyses, the discussion on AVN will differ; this will focus more on contrasts to CS and DHS as they have been stated to be the most commonly used IF treatments. Again, other interesting findings will also be discussed.

First, all arthroplasties performed better than IF. THAU was ranked as the best surgical treatment for displaced FNF at one-year follow-up in terms of the difference in QoL from pre- to post-operation, whereas IF was ranked last. Furthermore, Second, when comparing cemented to uncemented implants, hemiarthroplasties performed better when cemented compared to uncemented, although only BPHAC was significantly better than BPHAU. While there was a total of 0.241 standard deviations between BPHAC and BPHAU, this was very close to not being significant, as indicated by the 95%CI of 0.002 to 0.480. The difference between UPHAC and UPHAU however was much less, with 0.023 standard deviations between the two and no significant difference found.

The opposite was true for cementing in total hip arthroplasty; THAU performed better than THAC, albeit not significantly. The authors of the paper used in this study recommended against the use of THAU as it was associated with a substantially higher number of complications. When discussing the results for EQ-5D, they state that there was no difference. While both groups had the same scores at one year, they did not have equal baseline scores. Those with THAC had an average score of 0.9 at baseline, whereas those with THAU had 0.8, representing a 10% difference in the QoL estimate. As both ended up with an average score of 0.8, those with THAC in fact suffered a 10% loss in their scores, whereas those with THAU seem to have regained their pre-operative QoL. It was also not clear why the authors of this study reported the scores with only one decimal point; rounding to this degree is very misleading. As it will be discussed below, one issue with this study is the small sample size of certain treatment arms, and since THAU had only 34 participants. The superiority of THAU over other treatments from our results should therefore take this into consideration.

Third, results were conflicting when comparing bipolar and unipolar hemiarthroplasties. BPHAC was ranked better than UPHAC, while UPHAU was ranked better than BPHAU. Both contrasts were not significant. Fourth, all total hip arthroplasties performed better than hemiarthroplasties. When considering significance, THAU was ranked significantly better than BPHAC and BPHAU, while THAC was ranked significantly better than BPHAU.

The QoL results of this study are contrasted to those of the studies that reported non-EQ-5D indexes that could not be converted as comparison to studies reporting EQ-5D would be redundant. Comparisons for EQ-5D and mortality were limited to RCTs comparing surgical treatments for displaced FNF. When comparing SF-36 scores for total hip arthroplasty (cemented and uncemented) to hemiarthroplasties (bipolar and unipolar, cemented and uncemented) at one year, Macaulay and colleagues found that THA performed significantly better than HA for the bodily pain component of the SF36 (42.4 vs. 53.2, p=0.02). Authors recommend THA, however, only 40 patients were included [92]. Since our results show that both THA treatments performed better than all HA treatments, these results are supported. Avery and colleagues compared UPHAC to THAC (n=41 and 40, respectively) for SF-36 at 3yrs post-op and found no significant differences between the two. While THAC is ranked higher, our study does not show significance and is thus supported. El-Abed et al. compared DHS to

UPHAU (n=62 and 60, respectively) using the SF-36 measured at 3 years post-op and found that the latter led to a better QoL (p=0.002) [93]. This is also supported here.

Calder and colleagues compared DHS, UPHAC and BPHAC in terms of the NHP. They stratified their patients into two age groups. Group A was composed of patients 65 to 79 years and B was composed of individuals over 80 years. For group A showed that the DHS group has significantly worse scores than the hemiarthroplasties for the social index score (p=0.049), but not other components of the NHP. The UPHAC was nearly significantly better than DHS and BPHAC for the physical mobility (p=0.076) and energy (0.09) indexes of the NHP. While superiority of hemiarthroplasties is found in our study, the that of UPHC over BPAHC is not supported. Note however that these results were not significant, and that this was the younger (65-79 years) group and that only one study (of nine) had an average age in this range. For Group B, no significant differences were found but BPHAC was nearly significantly better than UPHAU for the pain index (p=0.065), which corroborates with our results. Note that the samples were relatively small (37, 39 and 34 for DHS, BPHAC and UPHAC, respectively for Group A, and 56 and 72 for BPHAC and UPHAC, respectively for Group B) and that follow-up was at 6 months only [94]. Lastly, Raia et al. compared BPHAC and UPHAC and found no significant differences in SF-36 scores at 1 year, which agrees with our results [95].

One study comparing UPHAU to IF was excluded as it only reported EQ-5D at two years and contributed significant inconsistency to the network [96]. Results from this investigation showed that UPHAU resulted in significantly greater loss of quality of life than IF. This is contradictory to the results presented here. Future studies into longer term QoL could be recommended to elucidate the matter. However, as both a ranked lowly, a recommendation for further studies will be avoided at this time.

The QoL NMA contributes to the literature in a few ways. First, this is the first QoL network meta-analysis comparing treatments for FNF. Second, this permitted a total of 21 contrasts to be calculated, of which 7 are significant. A total of 9 contrasts were initially available in the literature; this study therefore contributed a total of 12 contrasts. Third, this study produced a ranking of treatments that can be used by clinicians to optimize the treatment of displaced FNF. Fourth, the increased statistical power of using a network meta-analysis further validated several observations mentioned above already in the literature (ex. that cemented arthroplasties should be favored).

As previously described, if given treatments A and B, the OR is the odds of developing the outcome given that they received treatment A, divided by the odds of developing the outcome given that they received treatment B. An odds ratio of 1 indicates no difference between treatments, an OR < 1 indicates that treatment A is better than B, and an OR > 1 indicates that treatment B is better than A. For a OR of 0.5, treatment A has 50% lower, or half the odds of experiencing the outcome than treatment B. An OR of 2 means that treatment A has 100% higher, or twice the odds of experiencing the outcome than treatment B.

A proxy to statistical significance testing for the OR is the observation that the treatment effect 95%CI does not contain 1 (i.e. that the treatments have equal odds of being associated with the outcome of interest). Since none of the 95% CI of the treatment effects contain 1, the results of the mortality analysis are interpreted as there being equal odds of dying in each treatment arm.

As mentioned, the mortality analysis was restrained to only include data for displaced fractures and at one-year follow-up, with IF treatments grouped together. Sensitivity analyses showed that this produced similar results as when all follow-ups and fracture displacements were included and individual IF treatments analyzed. The rationale for this restriction was to facilitate comparisons and eventual combination of rankings. All analyses conducted showed no significant differences between treatments and did not have significant inconsistency. As there was no significant contrast, readers are cautioned that rankings need to be interpreted accordingly. There are a couple interesting findings to mention. First, IF performed better than all arthroplasties except for UPHAC. Second, all cemented implants were superior to their uncemented counterparts. Third, all unipolar hemiarthroplasties ranked higher than bipolar ones. Fourth, total hip arthroplasty performed better than all arthroplasties, except for UPHAC. Furthermore, while THAU is ranked best for QoL, it is ranked second-last before BPHAU for mortality. Similarly, although IF is ranked last for QoL, it has the second-best ranking for mortality. The contrast closest to being significant is UPHAC:UPHAU (OR=1.14, 95%CI=0.89 to 1.46), followed by BPHAU:UPHAC (OR=1.55, 95%CI = 0.88 to 2.73), then UPHAU:IF (OR=1.12, 95%CI=0.87 to 1.44).

While there was no significant inconsistency, the IF:UPHAC directly measured contrast contributed some inconsistency to its own summary treatment effect estimate (see Figure 3.11). As indicated by the grey square, the summary statistic for the IF:UPHAC (row) was informed by the direct IF:UPHAC (column) contrast. However, detaching of this contrast led to a reduction in inconsistency, implying that it contributed inconsistency when attached. Since it contributes inconsistency to itself, there must be a source of inconsistency within this design. This may be because two of the four studies that include this comparison included only patients in poor mental health, whereas the other studies did not exclude patients based on mental capacity. This is an example of heterogeneity that may have expressed itself as statistical inconsistency, yet not enough to be significant.

The 71 systematic reviews, meta-analyses and network meta-analyses identified in the literature search were reviewed for mortality contrasts. As supported by the current study, no contrasts were found to be significant.

This analysis contributes to the literature in various ways. This is the first proper mortality NMA conducted, contributing an additional 10 contrasts not previously made. It validates the existing literature on the matter with improved statistical power. It also produces a ranking of treatments that can be informative yet not conclusive (due to lack of significance) for physicians when choosing a treatment.

In contrast to the previous analyses, that for AVN was not restrained to new selection criteria. Papers included patients with mostly women with either displaced or undisplaced FNF and younger than previous studies. While the majority of studies are composed of mostly women, two (Lykke and Liu) have a minority of women. Results of this analysis therefore need to be put in this context.

Analyses for AVN found five significant contrasts, which are illustrated below along with rankings. CSIG was ranked best and was found to be significantly better that CS (OR=0.15, 95%CI=0.03 to 0.78) and CSC (OR=0.05, 95%CI=0.01 to 0.46). Implants with a dynamic component (SNP and DHS) performed well; SNP was ranked second best and DHS third best. However, DHS was the only that was significantly better than another treatment (DHS:CSC, OR=0.21, 95%CI 0.05 to 0.94). Hence, while SNP is ranked highly, it is not significantly better than any other treatment. CS performed significantly worse than Pins (OR=1.54, 95%CI=1.07 to 2.23]. The odds of experiencing AVN are therefore 154% greater when treated using CS than with Pins, with the true population effect being between 107% and 223%. The contrast CS:DHS was nearly significant with an OR = 1.67 and 95%CI = 0.95 to 2.93.

The largest OR was observed for the contrast between CSC and CSIG (OR=18.46 95%CI=2.19 to 155.71), which, along with all others including CSIG for that matter, also had large 95%CIs. This is most likely due to the small sample size of the CSIG group. Other treatments with small sample sizes are CSFG, CSC, and SNP, although these treatments were associated with smaller 95%CIs. One contrast that did have a satisfactory sample size is the CS:Pins.

Table 4.2: Ranking of treatments according to AVN ORs. Significant contrasts are indicated by brackets.

Treatment	P-score
CSIG	0.938
SNP	0.631
DHS	0.624

Pins	0.572
CSFG	0.496
Nail	0.437
CS	0.254
CSC	0.049

In a meta-analysis conducted in 1998 comparing screws to DHS and screws to pins (hooked and non-hooked) for the treatment of both displaced and nondisplaced FNF, Parker and Blundell found no significant differences in fracture healing complication rates and non-unions [97]. Another meta-analysis compared internal fixation devices in 2001 and again found no significant differences [98]. It is worth noting that the comparison between DHS and CS in terms of avascular necrosis rates was nearly significant (OR = 0.62, 95% CI = 0.38 to 1.01]. A Cochrane review conducted in 2011 comparing a total of 19 different pin and/or screw combinations found no differences in fracture healing, osteonecrosis, wound infection, pain scores, reoperation rate, use of walking aids, periprosthetic fracture or mortality [99].

This analysis contributes the following to the literature: it is the first NMA comparing treatments for FNF according to AVN, contributing 28 contrasts, 19 of which are new, and 5 are significant. Three significant contrasts were identified for the first time. It is the first time a significant difference is found between pins and cannulated screws. Furthermore, even with increased statistical power, the DHS:Pins contrast remains insignificant.

The poset analysis presented makes use of the results of the QoL and mortality results and needs to be interpreted accordingly. While the Hasse Diagram gives an overall ranking, the significance of the QoL and mortality analyses must be considered. As previously explained, for a treatment to be placed above and joined to another treatment in the diagram, the treatment needs to be superior for both outcomes analyzed. As expected, the cemented HA are ranked higher than their uncemented counterparts. Among cemented arthroplasties, THAC performed better than BPHAC but not UPHAC in both outcomes. Furthermore, IF was ranked below UPHAC for QoL and mortality. IF and BPHAU are at the lowest points in the diagram as they are worst in at least one outcome; they therefore cannot be better than other treatments for both outcomes.

If considering the poset results, it would therefore be advised that cement be used for arthroplasties. Furthermore, choice of an arthroplasty should be between UPHAC, THAC and THAU, not BPHAC. Although uncorroborated by significance testing, THAU was ranked second last for mortality. It was also associated with small sample sizes and large variance. These issues raise slight concerns that require that its endorsement be postponed. More studies are needed at this time. The use of cement is also recommended in various guidelines but the preference of a unipolar over bipolar hemiarthroplasty has not been stated an is therefore an original contribution to the literature.

The contributions of this analysis to the literature is that to the knowledge of the authors, it is the first poset analysis conducted in orthopaedics, further validating various national guideline recommendations (ex. cementing vs. uncemented). A clinical interpretation of combining QoL and mortality results can be presented as follows: if minimizing the risk of mortality is the priority, UPHAC should be used but following the necessary preventive treatments to maximize QoL. Unfortunately, one of the major complications of UPHAC, acetabular wear, is not easily prevented nor treated. Furthermore, the suggestion that UPHAC be used is based on the ranking of treatments and significance testing. If the latter is to be followed strictly, the recommendation would be for the attending surgeon to select between THAU, THAC and UPHAC. The NICE guideline offers a good basis for this selection: that THA be used in more active/fit individuals. Other treatments should be avoided as there is a more successful treatment than each of them. Again, there is a need to put this statement in context; it is a conclusion that has been made from data on displaced FNF in a mostly elderly, female European population. These results are likely to be different if another population was analyzed, for example youth with FNF. While the standard for treatment according to national guidelines is to use IF and arthroplasty for undisplaced fractures and displaced fractures, respectively. However, the studies included in this study only evaluate the elderly population and therefore should not be accepted as representing a younger population. Since conserving the femoral head is a priority, more research should be conducted comparing treatments for FNF in the non-elderly population. However, since most of these fractures occur in the elderly, there may be a lack of interest in studying another population. A potential solution could be to employ a random-effects network meta-regression analysis, with age as a covariate. This will unfortunately need many more studies, with a wide range of ages, to show an effect.

For this population, the AVN analysis gave some insight into the optimal treatment among IF options. It is worrisome as the literature states that CS and DHS are the most commonly used IF techniques since Pins were found to be significantly better than CS. DHS was the most successful treatment among currently used treatments. Although it was ranked higher than pins, there is no significant contrast between the two. The clinical interpretation of these results is as follows: for young individuals presenting with a FNF, the surgeon should choose between DHS and Pins to minimize the occurrence of AVN. Although recommendations for the use of CSIG is avoided here due to the small sample size and large estimate variance, future research into this treatment is recommended.

In addition to the sources of heterogeneity discussed above, there is a need to consider surgeon/hospital volume. Numerous studies have shown that these factors have an affect on patient outcomes; patients seen for a total hip replacement in higher surgeon/hospital volumes were more satisfied with their procedures and experienced less complications[100, 101]. This highlights that this variable could be informative in future NMAs and that it be potentially be used in a meta-regression type study to determine if it has an effect on treatment effect estimations.

Critique of the literature

To the knowledge of the authors of this paper, this is the only NMA conducted to date that compares FNF fracture for QoL and AVN. To date, there have been three network meta-analyses comparing surgical treatments for femoral neck fractures. The first, conducted by Liang and colleagues in 2015, compared treatments for FNF according to Harris Hip Scores. The second, by Mosseri and colleagues in 2016, compared treatments for FNF according to all-cause reoperation. The third, by Zhang et al. in 2017 looked at reoperation, mortality, infection, and dislocation.

Zhang and colleagues conducted an NMA to compare surgical interventions for displaced FNF in the elderly, which was published in Nature in October 2017 [102]. They compared treatments in terms of reoperation, mortality, dislocation and infection, and found a total of 40 studies and 6,141 participants. A few issues were identified while reviewing this paper.

The first ~~major~~ issue is the selection criteria. There is no indication as to whether papers retrieved need to report each of the listed outcomes, or any single one of them. This made it particularly difficult to review their work and could have been solved by simply indicating which paper reports which outcome. Furthermore, they do not mention follow-up times in their inclusion criteria and include papers with varying follow-up times. For example, Parker 2010 (4) is a study that has follow-up times of two to five years, while others (ex. Raia 2003) have follow-up times of three months and one year. Inclusion of papers with varying follow-up times could be acceptable if appropriate analyses were conducted and reported to ensure the combination of studies with different follow-up times was appropriate. However, this seems to not have been appropriate as they found significant inconsistency in their mortality analyses (BPHAC-THAC-IF loop; Inconsistency Factor = 0.75; p=0.021). While they do discuss potential sources of heterogeneity in their discussion, they do not mention follow-up times.

As highlighted in their work, some surgical operations for FNF have higher risks of reoperation and certain complications. The risk of these issues is not linear. For example, avascular necrosis present after one-year post-operation and require reoperation to be treated. As these surgeries are associated with a risk of death, the risk increases around the time of AVN diagnosis (i.e. after one year post-operatively). As many RCTS in this field have standard follow-up times of one year, two years and beyond, it would be recommended that they split their analyses into different follow-up times (ex. one year, two years, and so on).

Another major problem with their study is the search strategy. While the authors do include the basic databases needed for a systematic literature review (PubMed, EMBASE and Cochrane Library), they do not include Web of Science and the many RCT registries. They also do not include other sources, such as other systematic reviews and meta-analyses. The authors are praised for including at least one of their detailed search queries in supplementary materials. However, upon investigation, there are a couple issues that should be discussed. First, there are a few terms missing that should have been included. For example, they only included the term "femoral neck fractures" without any synonyms or subtypes (ex. cervical hip fracture). Second, they only included the population and randomized/RCT portions of the query and left out interventions.

These issues with their search strategy could be why they seem to have missed a few papers (ex. Avery and colleagues [103]). In addition to missing papers, they included duplicates; Blomfeldt 2007 and Hedbeck 2011 (46); and are duplicates. There are also some minor ones that are worth mentioning. There was only one bias assessor and all IF methods are grouped together; while this may be appropriate, a mention that analyses were conducted to support this would be beneficial.

As this study missed papers and included duplicates, the results are deemed to be biased. The papers that were missed did not contribute their data to the analyses, and those that were duplicated contributed too much. An overview of their results and its implications is therefore avoided here.

Mosseri and colleagues conducted an NMA in 2016 to compare treatments in terms of surgical reoperation [104]. They found a total of 27 RCTs that included 4,186 patients. There are a few strong points that are worth mentioning. First, they included other sources of studies (meta-analyses and RCT registries). Second, they did not group all IF together until they conducted analyses and consulted with specialists to support this. Not only did they conduct sensitivity analyses to show their groupings was appropriate, but they also consulted the literature and experts to substantiate this. Third, they put their results in the proper context; they state prior to making their conclusion that this is representative of a population that is "predominantly older women with displaced femoral neck fracture" rather than vaguely stating their results, as if they were true in any setting. Fourth, they mention that they reviewed the methodological and clinical characteristics of the studies and mention, even through they've shown there is no significant inconsistency, that the studies are comparable. Fifth, they referred to the PRISMA guidelines for writing NMAs and filled out the PRISMA form and included it as a supplementary document. Sixth, they included each of their search queries as supplementary material and they included synonyms to ensure they did not miss any studies.

Ultimately, this study shows that there is a significantly lower risk of revision surgery in patients treated with HA and THA when compared to screws, unthreaded cervical osteosynthesis, and plates. If an increase in the risk of revision can be interpreted as extrapolatable to an increase in the risk of poor QoL, these results agree with our results for QoL. While there are no major issues, a couple minor issues are worth mentioning. As stated by the authors, this is one of several clinically important outcomes; there remains a need for others to be investigated. Also, there is a need to consider cementing; they included studies comparing cemented to uncemented treatments, but only used this data for their first network graph, they did not conduct an NMA with these studies and there are therefore no pairwise contrasts for these treatments.

The third and last NMA was conducted by Liang and colleagues in 2015, who found a total of 15 studies with 1781 cases comparing treatments for FNF according to the Harris Hip Score [105]. In addition to the fact that they grouped IF treatments together without validating if this was proper, there were a few problems with their search strategy. First, their inclusion criteria states that “[studies] without significant comparison between surgeries” were excluded. It is not clear what is meant by this but if they excluded papers that did not find significant contrasts between treatments, this is not proper; all evidence should be included. Second, they did not include EMBASE, which is necessary according to various systematic literature review guidelines. Third, they describe two treatment groups as HA and artificial femoral head replacement; it is unclear which refers to unipolar or bipolar HA. Furthermore, their search query does not seem adequate as it is missing key synonyms.

Altogether, these issues may be why they only identified 15 studies. According to the results for the study being discussed here, a total of 40 studies were identified, albeit including duplicates, and some of these were missing from their sample (ex. Abdelkhalek [106] and Davison [107]). As papers were missing from their analyses, their results are not discussed here. It would be interesting to repeat this study ensuring no papers are missed and to use the SMD to enable the inclusion of papers reporting other hip scores, such as the Oxford Hip Score.

Limitations and solutions (next steps)

One limitation of this study is the small sample size of studies and participants. Analyses for EQ-5D and AVN have small sample sizes and there is especially a lack of studies investigating THAU for EQ-5D and mortality, and SNP, CSC, CSFG and CSIG for AVN. As evidence has suggested that THAU and THAC are superior treatments, additional research comparing these treatments to themselves or to other treatments would be ethical and offer more evidence for this sort of analysis to be repeated in the future. The study that included SNP was conducted in 1981, and there has since been no other study using it. In fact, SNP devices were replaced by DHS since screws offered a less invasive and more rigid means of stabilizing the fracture [108]. CSC also has a low number of participants and was found to have a higher, although insignificant, incidence of complications than CS [109]. Since the use of SNP has been discontinued, and CSC seems to be inferior than CS, the recommendation of more studies investigating these studies is avoided here. CSFG and CSIG also had low participant numbers. CSFG was not superior to CS in any outcomes measured in the included study, and recommendations for more research on this technique is therefore also avoided here [110].

On the other hand, results for CSIG suggest that this approach could be advantageous, especially for AVN. To be more specific, the treatment described in the study is a capsulotomy reduction and iliac crest autograft. According to the authors of the included study, this technique has the theoretical advantages of (i) enabling accurate anatomical reduction, which ensure bone-to-bone contact for improved bone healing and could liberate some obstructed retinacular vessels for improved blood flow to the femoral head (ii) offering stable internal fixation by enabling visualization during screw placement, offering capillaries support to grow into the ossiferous space, and creating conditions that are ideal for osteogenesis. They also refer to the induction osteogenesis, osteoconductive and angiogenic functions of bone grafting, and reconstruction of the blood supply and draining of the joint capsule during the capsulotomy. This technique seems to be promising and more research, with larger sample sizes, would be beneficial for future NMAs.

In addition to conducting more RCTs, inclusion of non-randomized studies could alleviate the low-sample size issues. This has been proposed by experts in the field of NMAs to be acceptable but that substantial precaution is needed [111]. This can be undertaken in future studies but was not feasible at this time. Furthermore, conference/association databases were not included, and the authors of ongoing studies were not contacted to gather preliminary data. As more papers were identified in our study than in meta-analyses or NMAs reviewed, these issues seem to not have had an effect. Nonetheless, a solution to this would be to include these sources in future studies. Also, some studies were excluded due to a lack of data in the papers. Authors could have been contacted but additional resources would have been needed. Rather, it is recommended that authors ensure they publish their work with adequate data; if they are reporting EQ-5D, they should include a table with actual average scores instead of simply including a graph for example (as in Blomfeldt and colleagues [61]).

As previously mentioned, only one study could be converted to EQ-5D. This was unfortunate since three studies reported Oxford Hip Scores, three reported SF-36, and one reported NHP. The HERC database for mapping studies from health indexes to EQ-5D can be of tremendous use to gather the largest number of studies possible but unfortunately, the data presented in the studies reporting non-EQ-5D QoL indexes did not report sufficient data for conversion. Furthermore, 42 papers (duplicates not considered) reported Harris Hip Scores. Since there is a validated algorithm to convert the Oxford Hip Score to EQ-5D, the Harris Hip Score could potentially be convertible to EQ-5D, which would increase our sample size considerably. A potential criticism of this study is that if the SMD was used, why not skip converting non-EQ-5D indexes to EQ-5D and simply include all QoL indexes in the SMD analysis? First, some studies that reported indexes that could be converted to EQ-5D were not truly QoL indicators and should therefore not be combined with the EQ-5D (ex. Oxford Hip Score, Pain scores). Others (ex. SF-12, SF-36, and NHP) evaluated QoL using multiple variables and do not give a single index that can be used in a SMD analysis. Therefore, converting studies to EQ-5D maximized the number of papers that could be included, while keeping the SMD representative of QoL.

There is also a need to recall that this study is limited to only analyzing EQ-5D, mortality and AVN, and that other important outcomes exist. There are a few key important outcomes when analyzing orthopedic treatments, such as mortality, infections, reoperation and hip function, yet with the resources available, a decision was made to focus on quality of life and mortality as they were deemed to give an overall picture of efficacy and safety. While an analysis of the QoL and mortality accomplish this, an analysis of other outcomes would be beneficial. For example, thromboembolism is another major complication that could be analyzed. AVN was chosen as the first complication to assess due to its associated severity and frequency in FNF patients. More studies can be conducted to give an exhaustive overview of the best treatments for each complication. This can then lead to further poset analyses that will take all complications into account, which will give a single ranking for the best treatment for FNF.

While some limitations may have affected the results of this study by limiting the number of studies retrieved or outcomes analyzed, there are also issues that made it substantially harder to conduct this study. These issues are often easy to solve and will be enumerated here to remind those conducting RCTs how to facilitate the work of those conducting meta-analyses and NMAs as it accepted as the next level of evidence quality. First, standard deviation or variance needs to be reported. Second, flowcharts showing patient numbers at randomization, surgical procedure and at each follow-up (with number of withdrawals and/or death and/or unknown) is needed. Third, the intention-to-treat principle should be used, and the exact method should be reported. Fourth, the exact treatments used in the studies need to be explained. It was burdensome to have to decipher which treatment was used in several studies; for example, many did not report whether cement was used. However, additional research was conducted and clarified the issue. For example, some authors may not have reported the use of cement, but used an Austin Moore hemiarthroplasty, which can only be used without cement. This approach was used in many cases. Fifth, an algorithm should be developed to convert HHS to EQ-5D as this would increase the number of studies that can be included in such studies. Sixth, when using EQ-5D in a RCT, reporting of which EQ-5D (3L or 5L) was used would be informative. Seventh, since the outcomes differ between displaced and undisplaced fractures, authors should stratify their analyses accordingly, or at the least state the distribution.

Another limitation to the present study is the lack of comparisons between internal fixation techniques for the EQ-5D analysis. While that for mortality gave no significant differences when each IF technique was included, and no inconsistency was found when they were joined into a single IF group, analysis of differences in QoL would have been informative, especially since significant contrasts were found for the AVN analysis. The rationale for not analyzing each IF treatment is that there was not a sufficient amount of comparisons available. More studies are need comparing IF treatments to each other for EQ-5D.

As more RCTs are surely to be conducted in coming years, there will be a need to update the NMAs reported here as papers are published. As previously explained, there are outcomes not analyzed here that would be beneficial to analyzed. Future directions of the work done here will be to analyze more outcomes (thromboembolism for example) and include future publications into the previous analyses; this study is the start of a series of NMAs on femoral neck fracture treatments.

Conclusion and summary

This study was composed of a series of NMAs that ranked treatments for FNF according to QoL, mortality and AVN. According to the results of the QoL and mortality NMAs and combined rankings, either THAU, THAC or UPHAC should be used as a treatment for displaced FNF in the elderly. While the former two were ranked poorly (sixth and third, respectively) for mortality, they were not significantly worse than other treatments. Cemented HA performed better than uncemented HA for QoL, and all cemented arthroplasties performed better than uncemented arthroplasties for mortality. Among cemented HAs, bipolar ranked higher than unipolar for QoL, but the opposite was true for mortality. The AVN NMA compared IF treatments and found CSIG to be ranked highest an CSC to be lowest. Pins were found to be significantly superior to CS but equivalent to DHS. This highlights that surgeons using CS should revaluate this approach and only use it with appropriate justification. More research is recommended on CSIG before it can be recommended as a replacement to pins or DHS as it was not found to be significantly better than them.

The results of this study should be interpreted as an additional tool that surgeons can use to facilitate their decision making; it is not intended to dictate which treatment to use but rather to validate or inform their choice. It is not only important to remember the context in which the data was collected (patient demographics, treatments used, etc.) but also that the outcomes analyzed here are among many others that should be considered when identifying which treatment to use. Nonetheless, this study contributes unique insights into which treatment is most likely to be best according to three important outcomes.

Appendix A

Table A.1: Web of Science (Core Collection) search query and results

ID	Search query	Number of results
#11	#10 AND #9 *DocType=All document types; Language=All languages;*	1,771
#10	TS= clinical trial* OR TS=research design OR TS=comparative stud* OR TS=evaluation stud* OR TS=controlled trial* OR TS=follow-up stud* OR TS=prospective stud* OR TS=random* OR TS=placebo* OR TS=(single blind*) OR TS=(double blind*) *DocType=All document types; Language=All languages;*	3,781,400
#9	#8 NOT #5 *DocType=All document types; Language=All languages;*	4,796
#8	#7 AND #6 *DocType=All document types; Language=All languages;*	6,195
#7	#4 OR #3 *DocType=All document types; Language=All languages;*	7,069,626
#6	#2 OR #1 *DocType=All document types; Language=All languages;*	11,896
#5	TI = (Bisphosphonat* OR bisphosphate* OR diphosphonate* OR diphosphate* OR Denosumab* OR Teriparatid* OR Clopidrog* OR clopidog* OR estradiol* OR hormone* OR vitamin* OR estrogen* OR an?esthesi* OR calcitriol* OR calcitonin* OR Tranexam* OR Hemocoagulase* OR Anticoagulant* OR Metformin* OR Biochemical* OR Ultrasound* OR Glycerin enema* OR Pamidron* OR Neridron* OR Olpadron* OR Alendron* OR Ibandron* OR Risedron* OR Zoledron* OR etidron* OR strontium OR mineral OR tanezumab OR inhibitor OR abaloparatide OR fluoride OR nandrolone OR bazedoxifene OR odanacatib OR exercise* OR warfarin OR rat* OR mice OR monkey* OR dog* OR intrauterine OR vagin* OR uterus* OR endometri* OR animal* OR cadaver OR "hip protector" OR protein* OR oral* OR raloxifene OR an?emi* OR glucocorticoid* OR tomograph* OR blood OR serum) *DocType=All document types; Language=All languages;*	5,665,064
#4	TI = (surger* OR operati* OR therap* OR procedure* OR implant* OR prothetic* OR prosthes* OR nail* OR screw* OR plate* OR pin* OR fixation OR fixator* OR "bed rest" OR hemi?arthroplast* OR arthroplast* OR replacement*) *DocType=All document types; Language=All languages;*	2,064,707

#3	TS = (surger* OR operati* OR therap* OR procedure* OR implant* OR prothetic* OR prosthes* OR nail* OR screw* OR plate* OR pin* OR fixation OR fixator* OR "bed rest" OR hemi?arthroplast* OR arthroplast* OR replacement*) *DocType=All document types; Language=All languages;*	7,069,626
#2	TI = ("femoral neck" OR "neck of femur" OR "femur neck" OR intracapsular OR collum OR subcapital OR basicervical OR basocervical OR transcervical) AND TI = (fracture* OR break* OR broken*) *DocType=All document types; Language=All languages;*	3,731
#1	TS = ("femoral neck" OR "neck of femur" OR "femur neck" OR intracapsular OR collum OR subcapital OR basicervical OR basocervical OR transcervical) AND TS = (fracture* OR break* OR broken*) *DocType=All document types; Language=All languages;*	11,896

Table A.2: Cochrane search query and results:

ID	Search	Hits
#54	#53 not (bisphosphonates or bisphosphate or diphosphonate or diphosphate or pamidron* or "pamidronic acid" or neridron* or "neridronic acid" or olpadron* or "olpadronic acid" or alendron* or "alendronic acid" or ibandron* or "ibandronic acid" or risedron* or "risedronic acid" or zoledron* or "zoledronic acid" or teriparatide or clopidrog* or clopidog* or tranexamic or hemocoagulase or anticoagulants or heparin or metformin or aspirin or enema or estrogen or estradiol or hormone or vitamin or exercise or anesthesia or calcitriol or calcitonin):ti in Cochrane Reviews (Reviews and Protocols) and Trials	857
#53	#18 and #52	1280
#52	#50 or #51	677514
#51	#32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 in Cochrane Reviews (Reviews and Protocols), Other Reviews and Trials	583595
#50	#19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31	297377
#49	rest	18542
#48	implant	7094

#47	prosthetic	2140
#46	prosthesis	10578
#45	replacement	22351
#44	arthroplasty	7803
#43	hemiarthroplasty	336
#42	plate	2440
#41	nail	1605
#40	screw	1825
#39	fixator	274
#38	fixation	6145
#37	technique	42042
#36	surgery	161598
#35	operative procedure	6819
#34	general surgery	22294
#33	operation	23788
#32	treatment	488061
#31	MeSH descriptor: [Bed Rest] explode all trees	437
#30	MeSH descriptor: [Internal Fixators] explode all trees	1581
#29	MeSH descriptor: [Bone Plates] explode all trees	529
#28	MeSH descriptor: [Bone Screws] explode all trees	736
#27	MeSH descriptor: [Bone Nails] explode all trees	427
#26	MeSH descriptor: [Fracture Fixation] explode all trees	1619
#25	MeSH descriptor: [Prostheses and Implants] explode all trees	17361
#24	MeSH descriptor: [Arthroplasty, Replacement, Hip] explode all trees	1942
#23	MeSH descriptor: [Hemiarthroplasty] explode all trees	36
#22	MeSH descriptor: [Arthroplasty, Replacement, Hip] explode all trees	1942
#21	MeSH descriptor: [Orthopedic Procedures] explode all trees	12076
#20	MeSH descriptor: [Therapeutics] explode all trees	281963
#19	MeSH descriptor: [General Surgery] explode all trees	365
#18	#4 or #17	1456
#17	#15 and #16 in Cochrane Reviews (Reviews and Protocols), Other Reviews and Trials	1311
#16	#5 or #6 or #7 or #8 or #9 or #10 or #11	3293
#15	#12 or #13 or #14 in Cochrane Reviews (Reviews and Protocols) and Trials	13542
#14	broken	2344
#13	break	2259

#12	fracture	11060
#11	transcervical	315
#10	bas*cervical	2
#9	collum	50
#8	intracapsular	317
#7	neck of femur	180
#6	femur neck	1202
#5	femoral neck in Cochrane Reviews (Reviews and Protocols) and Trials	2349
#4	#1 or (#2 and #3)	437
#3	MeSH descriptor: [Fractures, Bone] explode all trees	4959
#2	MeSH descriptor: [Femur Neck] explode all trees	432
#1	MeSH descriptor: [Femoral Neck Fractures] explode all trees	343

Table A.3: Medline search query and results

1	exp Femoral Neck Fractures/	8238
2	exp Femur Neck/	6431
3	exp fractures, bone/	166491
4	femoral neck.ti,ab.	16366
5	Femur neck*.ti,ab.	1157
6	Collum*.ti,ab.	407
7	Neck of femur*.ti,ab.	1003
8	Intracapsular*.ti,ab.	3108
9	Subcapital*.ti,ab.	756
10	Basicervical*.ti,ab.	68
11	Basocervical.ti,ab.	7
12	Transcervical*.ti,ab.	2732
13	2 and 3	2158

14	1 or 13	9307
15	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12	24345
16	14 or 15	27446
17	exp *general surgery/	27757
18	exp *therapeutics/	2048273
19	exp *Orthopedic Procedures/	182817
20	exp *arthroplasty/	45346
21	exp *hemiarthroplasty/	425
22	exp *prosthesis implantation/	93879
23	exp *"prostheses and implants"/	321858
24	exp *fracture fixation/	38088
25	exp *fracture fixation, internal/	26700
26	exp *bone nails/	5377
27	exp *bone screws/	10229
28	exp *bone plates/	7768
29	exp *nails/	4868
30	exp *internal fixators/	26153
31	exp *bed rest/	1820
32	exp *traction/	3061
33	exp *Treatment Outcome/	7067
34	Treatment*.ti,ab.	3815678
35	Operati*.ti,ab.	755046
36	General surger*.ti,ab.	8607
37	Operative procedure*.ti,ab.	11674

38	Surger*.ti,ab.	969750
39	Procedure*.ti,ab.	966253
40	Surgical*.ti,ab.	874458
41	Ghost*.ti,ab.	8385
42	Placebo*.ti,ab.	195480
43	Sham*.ti,ab.	80362
44	Technique*.ti,ab.	1307797
45	Fixation*.ti,ab.	125707
46	Fixator*.ti,ab.	5929
47	Open reduction*.ti,ab.	9632
48	Screw*.ti,ab.	36708
49	Nail*.ti,ab.	28042
50	Plate*.ti,ab.	377794
51	Hemiarthroplast*.ti,ab.	2501
52	Hemi-arthroplast*.ti,ab.	138
53	Arthroplast*.ti,ab.	50129
54	Total hip replacement*.ti,ab.	8356
55	Hip replacement*.ti,ab.	11109
56	Replacement*.ti,ab.	223393
57	Prosthe*.ti,ab.	109458
58	Implant*.ti,ab.	341053
59	Rest*.ti,ab.	978485
60	Traction*.ti,ab.	17043
61	17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33	2507303

62	34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60	7893061
63	61 or 62	9149161
64	16 and 63	17862
65	exp clinical trial as topic/	313324
66	exp random allocation/	92731
67	Clinical.ti,ab.	2955146
68	trial.ti,ab.	498052
69	clinical.ti,ab.	2955146
70	trials.ti,ab.	455594
71	clinical trial.pt.	521540
72	random.ti,ab.	223179
73	exp Therapeutic Uses/	4845389
74	65 or 66	390539
75	67 and 68	219891
76	69 and 70	272589
77	71 or 72 or 73 or 75 or 76	5528484
78	74 or 77	5652982
79	64 and 78	3093
80	exp animal/	21286911
81	exp human/	16882076
82	bisphosph*.ti.	10569
83	denosumab*.ti.	865
84	teriparatid*.ti.	704

85	clopidrogel*.ti.	6
86	tranexamic*.ti.	1757
87	Hemocoagulase*.ti.	37
88	Anticoagulant*.ti.	18317
89	warfarin*.ti.	8327
90	metformin*.ti.	7814
91	biochemical*.ti.	65162
92	ultrasound*.ti.	74940
93	"glycerin enema*".ti.	11
94	pamidron*.ti.	991
95	neridron*.ti.	52
96	olpadron*.ti.	29
97	alendron*.ti.	2029
98	ibandron*.ti.	466
99	risedron*.ti.	656
100	zoledron*.ti.	2043
101	etidron*.ti.	500
102	calcitriol*.ti.	1460
103	retinol*.ti.	4071
104	steroid*.ti.	68420
105	androgen*.ti.	29131
106	hormon*.ti.	185277
107	raloxifen*.ti.	1369
108	mutat*.ti.	154798

109	protein*.ti.	752282
110	glucocorticoid*.ti.	23629
111	rat*.ti.	1061577
112	mice*.ti.	277696
113	cyclosporin*.ti.	20830
114	monkey*.ti.	45302
115	80 not 81	4404835
116	82 or 83 or 84 or 85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100 or 101 or 102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114	2648449
117	79 not 115 not 116	2219

Table A.4: Embase search query and results

ID	Search query	Results
1	exp femoral neck fracture/	439
2	exp femoral neck/	944
3	exp fracture/	270353
4	Femoral neck*.ti,ab.	22994
5	Femur neck*.ti,ab.	1404
6	Collum*.ti,ab.	618
7	Neck of femur*.ti,ab.	1708
8	Intracapsular*.ti,ab.	4146
9	Subcapital*.ti,ab.	1004
10	Basicervical*.ti,ab.	90
11	Basocervical.ti,ab.	13
12	Transcervical*.ti,ab.	3473
13	Fracture*.ti,ab.	273541
14	Break*.ti,ab.	245662
15	2 and 3	376

16	1 or 15	776
17	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12	33692
18	13 or 14	515251
19	17 and 18	16422
20	16 or 19	16620
21	exp hip surgery/	23414
22	exp Orthopaedic surgery/	416736
23	exp Hip arthroplasty/	18375
24	exp Hip hemiarthroplasty/	472
25	exp Implantation/	122258
26	exp fracture fixation/	81459
27	Treatment*.ti,ab.	5252546
28	Operati*.ti,ab.	1104196
29	General surger*.ti,ab.	11678
30	Operative procedure*.ti,ab.	17932
31	Surger*.ti,ab.	1340113
32	Procedure*.ti,ab.	1318398
33	Surgical*.ti,ab.	1189227
34	Ghost*.ti,ab.	9350
35	Placebo*.ti,ab.	258983
36	Sham*.ti,ab.	108237
37	Technique*.ti,ab.	1669908
38	Fixation*.ti,ab.	157396
39	Fixator*.ti,ab.	6567
40	Open reduction*.ti,ab.	11319
41	Screw*.ti,ab.	42755
42	Nail*.ti,ab.	39559
43	Plate*.ti,ab.	46872
44	Hemiarthroplast*.ti,ab.	2812
45	Hemi-arthroplast*.ti,ab.	157
46	Arthroplast*.ti,ab.	58323
47	Total hip replacement*.ti,ab.	10393
48	Hip replacement*.ti,ab.	14267
49	Replacement*.ti,ab.	291354
50	Prosthe*.ti,ab.	132554
51	Implant*.ti,ab.	451634

52	Rest*.ti,ab.	1243843
53	Traction*.ti,ab.	25194
54	21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53	10503649
55	20 and 54	11715
56	exp clinical trial/	1211105
57	Dt.fs.	3393064
58	Clinical.ti,ab.	4139029
59	Trial.ti,ab.	68103
60	Radom*.mp.	265
61	56 or 57	4042314
62	58 and 59	302164
63	60 or 62	302411
64	61 or 63	4131444
65	55 and 64	2075
66	Bisphosphonate*.ti.	7715
67	Denosumab.ti.	1643
68	Teriparatide.ti.	1101
69	Clopidrogel.ti.	14
70	Hemocoagulase*.ti.	47
71	Anticoagulant*.ti.	22869
72	Warfarin.ti.	11318
73	Metformin.ti.	12044
74	Biochemical*.ti.	78906
75	Ultrasound*.ti.	98847
76	Glycerin enema*.ti.	13
77	Neridronate.ti.	68
78	Olpadronate.ti.	32
79	Alendronate.ti.	2676
80	Ibandronate.ti.	685
81	Risedronate.ti.	898
82	Zoledronate.ti.	467

83	heparin.ti.	35348
84	vitamin*.ti.	117298
85	66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84	386324
86	65 not 85	1599

Appendix B

Filter to import from Clinicaltrials.gov:

http://endnote.com/downloads/filter/clinicaltrials

Filter for import from Cochrane Library:

http://endnote.com/downloads/filter/cochrane-library

Filter to import from WHO:

http://endnote.com/downloads/filter/who-international-clinical-trials-registry-platform-ictrp

Filter for RCTs in WOS:

http://guides.library.ualberta.ca/c.php?g=248586&p=1655962

Filter for RCTs in Scopus:

http://libguides.nus.edu.sg/c.php?g=145717&p=2470589

Appendix C

Table A.5: Cochrane bias assessment tool used in this study (cite)

RANDOM SEQUENCE GENERATION

<table>
<tr><td colspan="2">Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.</td></tr>
<tr><td>Criteria for a judgement of ‘Low risk’ of bias.</td><td>The investigators describe a random component in the sequence generation process such as:
• Referring to a random number table;
• Using a computer random number generator;
• Coin tossing;
• Shuffling cards or envelopes;
• Throwing dice;
• Drawing of lots;
• Minimization*.

*Minimization may be implemented without a random element, and this is considered to be equivalent to being random.</td></tr>
<tr><td>Criteria for the judgement of ‘High risk’ of bias.</td><td>The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example:
• Sequence generated by odd or even date of birth;
• Sequence generated by some rule based on date (or day) of admission;
• Sequence generated by some rule based on hospital or clinic record number.

Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgement or some method of non-random categorization of participants, for example:
• Allocation by judgement of the clinician;
• Allocation by preference of the participant;
• Allocation based on the results of a laboratory test or a series of tests;
• Allocation by availability of the intervention.</td></tr>
<tr><td>Criteria for the judgement of ‘Unclear risk’ of bias.</td><td>Insufficient information about the sequence generation process to permit judgement of ‘Low risk’ or ‘High risk’.</td></tr>
<tr><td colspan="2">ALLOCATION CONCEALMENT
Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment.</td></tr>
<tr><td>Criteria for a judgement of ‘Low risk’ of bias.</td><td>Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:</td></tr>
</table>

	• Central allocation (including telephone, web-based and pharmacy-controlled randomization); • Sequentially numbered drug containers of identical appearance; • Sequentially numbered, opaque, sealed envelopes.
Criteria for the judgement of ‘High risk’ of bias.	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on: • Using an open random allocation schedule (e.g. a list of random numbers); • Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or nonopaque or not sequentially numbered); • Alternation or rotation; • Date of birth; • Case record number; • Any other explicitly unconcealed procedure.
Criteria for the judgement of ‘Unclear risk’ of bias.	Insufficient information to permit judgement of ‘Low risk’ or ‘High risk’. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.

BLINDING OF PARTICIPANTS AND PERSONNEL

Performance bias due to knowledge of the allocated interventions by participants and personnel during the study.

Criteria for a judgement of ‘Low risk’ of bias.	Any one of the following: • No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding; • Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.
Criteria for the judgement of ‘High risk’ of bias.	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.
Criteria for the judgement of ‘Unclear risk’ of bias.	Any one of the following: • Insufficient information to permit judgement of ‘Low risk’ or ‘High risk’; • The study did not address this outcome.

BLINDING OF OUTCOME ASSESSMENT

Detection bias due to knowledge of the allocated interventions by outcome assessors.

Criteria for a judgement of 'Low risk' of bias.	Any one of the following: • No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding; • Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding; • Blinding of outcome assessment, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk'; • The study did not address this outcome.
INCOMPLETE OUTCOME DATA **Attrition bias due to amount, nature or handling of incomplete outcome data.**	
Criteria for a judgement of 'Low risk' of bias.	Any one of the following: • No missing outcome data; • Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); • Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size; • Missing data have been imputed using appropriate methods.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size;

	• 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization; • Potentially inappropriate application of simple imputation.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided); • The study did not address this outcome.
SELECTIVE REPORTING **Reporting bias due to selective outcome reporting.**	
Criteria for a judgement of 'Low risk' of bias.	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • Not all of the study's pre-specified primary outcomes have been reported; • One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; • One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); • One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; • The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk' or 'High risk'. It is likely that the majority of studies will fall into this category.
OTHER BIAS **Bias due to problems not covered elsewhere in the table.**	
Criteria for a judgement of 'Low risk' of bias.	The study appears to be free of other sources of bias.
Criteria for the judgement of 'High risk' of bias.	There is at least one important risk of bias. For example, the study: • Had a potential source of bias related to the specific study design used; or • Has been claimed to have been fraudulent; or

	• Had some other problem.
Criteria for the judgement of 'Unclear risk' of bias.	There may be a risk of bias, but there is either: • Insufficient information to assess whether an important risk of bias exists; or • Insufficient rationale or evidence that an identified problem will introduce bias.

Appendix D

Sf-12 conversion to EQ-5D algorithm:

EQ-5D Index = 0.8469 + (PCS12 – 49.9) × 0.01261 + (MCS12 – 51.5) ×0.00759 – (PCS12 – 49.9)2 × 0.00009 – (MCS12– 51.5)2 × 0.00015 – (PCS12 – 49.9) × (MCS12 – 51.5) × 0.00015.

Bibliography

1. Wehren LE, Magaziner J: **Hip fracture: risk factors and outcomes**. *Curr Osteoporos Rep* 2003, **1**(2):78-85.
2. Marks R: **Hip fracture epidemiological trends, outcomes, and risk factors, 1970–2009**. *International Journal of General Medicine* 2010, **3**:1-17.
3. Fitzpatrick LA: **Secondary causes of osteoporosis**. *Mayo Clin Proc* 2002, **77**(5):453-468.
4. Leslie WD, O'Donnell S, Lagace C, Walsh P, Bancej C, Jean S, Siminoski K, Kaiser S, Kendler DL, Jaglal S: **Population-based Canadian hip fracture rates with international comparisons**. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 2010, **21**(8):1317-1322.
5. Leslie WD, O'Donnell S, Jean S, Lagace C, Walsh P, Bancej C, Morin S, Hanley DA, Papaioannou A: **Trends in hip fracture rates in Canada**. *Jama* 2009, **302**(8):883-889.
6. Hopkins RB, Pullenayegum E, Goeree R, Adachi JD, Papaioannou A, Leslie WD, Tarride JE, Thabane L: **Estimation of the lifetime risk of hip fracture for women and men in Canada**. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 2012, **23**(3):921-927.
7. Wiktorowicz ME, Goeree R, Papaioannou A, Adachi JD, Papadimitropoulos E: **Economic Implications of Hip Fracture: Helath Service Use, Institutional Care and Cost in Canada**. *Osteoporosis International* 2001, **12**:271-278.
8. Eisman JA, Bogoch ER, Dell R, Harrington JT, McKinney RE, Jr., McLellan A, Mitchell PJ, Silverman S, Singleton R, Siris E: **Making the first fracture the last fracture: ASBMR task force report on secondary fracture prevention**. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* 2012, **27**(10):2039-2046.
9. Manninger J, Bosch U, Cserhati P, Fekete K, Kazar G: **Internal fixation of femoral neck fractures**. Austria: Springer Wien New York 2007.
10. Bishop JA, Braun BJ, Buza J, Davidovitch R, Duckworth AD, Dweck E, Egol KA, Ekkernkamp A, Gales J, Ganta A *et al*: **Proximal Femur Fractures: An Evidence-Based Approach to Evaluation and Management** Springer International Publishing 2018.
11. Forsh DA, Ferguson TA: **Contemporary management of femoral neck fractures: the young and the old**. *Current reviews in musculoskeletal medicine* 2012, **5**(3):214-221.
12. Sheehan SE, Shyu JY, Weaver MJ, Sodickson AD, Khurana B: **Proximal Femoral Fractures: What the Orthopedic Surgeon Wants to Know**. *Radiographics* 2015, **35**(5):1563-1584.
13. Lein T, Bula P, Jeffries J, Engler K, Bonnaire F: **Fractures of the femoral neck**. *Acta Chir Orthop Traumatol Cech* 2011, **78**(1):10-19.
14. Koval K, Zuckerman J: **Hip Fractures: A Practical Guide to Management**. New York, NY: Springer; 2000.
15. Zlowodzki M, Bhandari M, Keel M, Hanson BP, Schemitsch E: **Perception of Garden's classification for femoral neck fractures: an international survey of 298 orthopaedic trauma surgeons**. *Archives of orthopaedic and traumatic surgery Archiv fur orthopadische und Unfall-Chirurgie* 2005, **125**(7):503-505.
16. Mackey DC, Lui LY, Cawthon PM, Bauer DC, Nevitt MC, Cauley JA, Hillier TA, Lewis CE, Barrett-Connor E, Cummings SR: **High-Trauma Fractures and Low Bone Mineral Density in Older Women and Men**. *Journal of the American Medical Association* 2007, **298**(20):2381-2388.

17. Webb L: **Proximal femoral fractures**. *Journal of the Southern Orthopaedics Association* 2002, **11**(4):203-212.
18. Alho A: **Concurrent ipsilateral fractures of the hip and shaft of the femur. A systemic review of 722 cases.** . *Annales chirurgiae et gynaecologiae* 1997, **86**(4):326-336.
19. Peljovich AE, Patterson BM: **Ipsilateral femoral neck and shaft fractures**. *Journal of the American Orthopaedic Surgeons* 1998, **6**(2):106-113.
20. Marks R, Allegrante JP, Ronald MacKenzie C, Lane JM: **Hip fractures among the elderly: causes, consequences and control**. *Ageing Research Reviews* 2003, **2**(1):57-93.
21. Parker MJ: **Fractures of the neck of the femur**. *Trauma* 2008, **10**:43-53.
22. Fernandez MA, Griffin XL, Costa ML: **Management of hip fracture**. *Br Med Bull* 2015, **115**(1):165-172.
23. Bateman L, Vuppala S, Porada P, Carter W, Baijnath C, Burman K, Lee R, Hargus J: **Medical Management in the Acute Hip Fracture Patient: A Comprehensive Review for the Internist**. *The Ochsner Journal* 2012, **12**(2):101-110.
24. **Hip Fracture: Management**: National Institute for Health Care Excellence; 2011.
25. McGlasson R, Zellermeyer V, MacDonald V, Spafford D, McMullan JL, Beaupre L, Gould JS, Saryeddine T, Scott V: **National Hip Fracture Toolkit**: Bone and Joint Decade Canada; 2011.
26. Van Waesberghe J, Stevanovic A, Rossaint R, Coburn M: **General vs. neuraxial anaesthesia in hip fracture patients: a systematic review and meta-analysis**. *BMC Anesthesiol* 2017, **17**:87.
27. **Mangement of Hip Fractures in the Elderly: Evidence-based Clinical Practice Guidelines**: American Academy of Orthopaedic Surgeons 2014.
28. Ly TV, Swiontkowski MF: **Management of femoral neck fractures in young adults**. *Indian J Orthop* 2008, **42**(1):3-12.
29. **Fracture neck of femur – Still an unsolved issue**. *J Orthop* 2016, **13**(1):A1-A3.
30. Roshan A, Ram S: **The Neglected Femoral Neck Fracture in Young Adults: Review of a Challenging Problem**. *Clinical Medicine & Research* 2008, **6**(1):33-39.
31. Sandhu HS, Dhillon MS, Jain AK: **Femoral neck fractures**. *Indian J Orthop* 2008, **42**(1):1-2.
32. Hoffman R, Haas NP: **Fracture of the femoral neck (31-B)**. In: *AO Principles of Fracture Management.* edn. New York AO Publishing 2000.
33. Tronzo RG: **Hip nails for all occasions**. *The Orthopedic clinics of North America* 1974, **5**(3):479-491.
34. Rogmark C, Carlsson A, Johnell O, Sernbo I: **Arthroplasty Led to Fewer Failures and More Complications Than Did Internal Fixation for Displaced Fractures of the Femoral Neck**. *JBJS* 2002, **84**(11):2108.
35. Mittal R, Banerjee S: **Proximal femoral fractures: Principles of management and review of literature**. *Journal of clinical orthopaedics and trauma* 2012, **3**(1):15-23.
36. Bartonicek J, Rammelt S: **The history of internal fixation of proximal femur fractures Ernst Pohl-the genius behind**. *Int Orthop* 2014, **38**(11):2421-2426.
37. Knight SR, Aujla R, Biswas SP: **Total Hip Arthroplasty - over 100 years of operative history**. *Orthopedic Reviews* 2011, **3**(2):e16.
38. Marya SK, Thukral R, Singh C: **Prosthetic replacement in femoral neck fracture in the elderly: results and review of the literature**. *Indian J Orthop* 2008, **42**(1):61-67.
39. Hedbeck C, Enocson A, Lapidus G, Blomfeldt R, Törnkvist H, Ponzer S, Tidermark J: **Comparison of bipolar hemiarthroplasty with total hip arthroplasty for displaced femoral neck fractures: a concise four-year follow-up of a randomized trial**. In: *The Journal of bone and joint surgery American volume.* vol. \93; 2011: 445-450.

40. López-López JA, Humphriss RL, Beswick AD, Thom HHZ, Hunt LP, Burston A, Fawsitt CG, Hollingworth W, Higgins JPT, Welton NJ *et al*: **Choice of implant combinations in total hip replacement: systematic review and network meta-analysis**. 2017, **359**.
41. Moretti VM, Post ZD: **Surgical Approaches for Total Hip Arthroplasty**. *Indian J Orthop* 2017, **51**(4):368-376.
42. Horne PH, Olson SA: **Direct anterior approach for total hip arthroplasty using the fracture table**. *Current reviews in musculoskeletal medicine* 2011, **4**(3):139-145.
43. Skoldenberg O, Ekman A, Salemyr M, Boden H: **Reduced dislocation rate after hip arthroplasty for femoral neck fractures when changing from posterolateral to anterolateral approach**. *Acta Orthop* 2010, **81**(5):583-587.
44. Biber R, Brem M, Singler K, Moellers M, Sieber C, Bail HJ: **Dorsal versus transgluteal approach for hip hemiarthroplasty: an analysis of early complications in seven hundred and four consecutive cases**. *Int Orthop* 2012, **36**(11):2219-2223.
45. Neogi DS, Kumar KVA, Trikha V, Yadav CS: **Ipsilateral femoral neck and trochanter fracture**. *Indian Journal of Orthopedics* 2011, **45**(1):82-86.
46. Dhar SA, Mir MR, Butt MF, Farooq M, Ali MF: **Osteosynthesis for a T-shaped fracture of the femoral neck and trochanter: a case report**. *Journal of Orthopaedic Surgery* 2008, **16**(2):257-259.
47. Butler M, Forte M, Kane RL, Joglekar S, Duval SJ, Swiontkowski M, Wilt T: **Treatment of Common Hip Fractures. Evidence Report/Technology Assessment No. 184**. In. Edited by 1. ME-bPCuCNH. Rockville, MD. : Agency for Healthcare Research and Quality 2009.
48. Leslie WD, O'Donnell S, Lagacé C, Walsh P, Bancej C, Jean S, Siminoski K, Kaiser S, Kendler DL, Jaglal S: **Population-based Canadian hip fracture rates with international comparisons**. *Osteoporosis International* 2010, **21**:1317-1322.
49. Cauley JA: **Defining ethnic and racial differences in osteoporosis and fragility fractures**. *Clin Orthop Relat Res* 2011, **469**(7):1891-1899.
50. Hailer NP, Garland A, Rogmark C, Garellick G, Karrholm J: **Early mortality and morbidity after total hip arthroplasty in patients with femoral neck fracture**. *Acta Orthop* 2016, **87**(6):560-566.
51. Kilci O, Un C, Sacan O, Gamli M, Baskan S, Baydar M, Ozkurt B: **Postoperative Mortality after Hip Fracture Surgery: A 3 Years Follow Up**. *PLoS One* 2016, **11**(10):e0162097.
52. Kim SD, Park SJ, Lee DH, Jee DL: **Risk factors of morbidity and mortality following hip fracture surgery**. *Korean J Anesthesiol* 2013, **64**(6):505-510.
53. Lawrence VA, Hilsenbeck SG, Noveck H, Poses RM, Carson JL: **Medical complications and outcomes after hip fracture repair**. *Arch Intern Med* 2002, **162**(18):2053-2057.
54. Poh KS, Lingaraj K: **Complications and their risk factors following hip fracture surgery**. *Journal of orthopaedic surgery (Hong Kong)* 2013, **21**(2):154-157.
55. Centre NCG: **The management of hip fractures in adults**. In. Edited by Centre NCG. London; 2011.
56. Carpintero P, Caeiro JR, Carpintero R, Morales A, Silva S, Mesa M: **Complications of hip fractures: A review**. *World journal of orthopedics* 2014, **5**(4):402-411.
57. Salkeld G, Cameron ID, Cumming RG, Easter S, Seymour J, Kurrle SE, Quine S: **Quality of life related to fear of falling and hip fracture in older women: a time trade off study** *British Medical Journal* 2000, **320**:341-346.
58. Parsons N, Griffin XL, Achten J, Costa ML: **Outcome assessment after hip fracture: is EQ-5D the answer?** *Bone Joint Res* 2014, **3**(3):69-75.

59. Dakin H, Abel L, Burns R, Yang Y: **Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement**. *Health Qual Life Outcomes* 2018, **16**(1):31.
60. Tonin FS, Rotta I, Mendes AM, Pontarolo R: **Network meta-analysis: a technique to gather evidence from direct and indirect comparisons**. *Pharmacy Practice* 2017, **15**(1):943.
61. **Office of the Surgeon General (US). Bone Health and Osteoporosis: A Report of the Surgeon General. Rockville (MD): Office of the Surgeon General (US); 2004. 5, The Burden of Bone Disease. Available from: https://www.ncbi.nlm.nih.gov/books/NBK45502/**.
62. Laws A, Kendall R, Hawkins N: **A Comparison of National Guidelines for Network Meta-Analysis**. *Value in Health* 2014, **17**(5):642-654.
63. **Cochrane Handbook for Systematic Reviews of Interventions** Version 5.1.0 [updated March 2011]. edn: The Cochrane Collaboration; 2011.
64. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG: **Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement**. *PLOS Medicine* 2009, **6**(7):e1000097.
65. Lefebvre C, Manheimer E, Glanville J: **Search for studies** In: *Cochrane Handbook for Systematic Reviews of Interventions* 5.1.0 edn.: The Cochrane Collaboration; 2011.
66. Dakin H., Burns R., Y. Y. In.; 2016.
67. Guenhan BK: **nmaINLA** In.: CRAN.R; 2017.
68. Rucker G, Schwarzer G, Krahn U, König J: **Network Meta-Analysis using Frequentist Methods**. In.: CRAN.R; 2017.
69. Higgins JP, Thompson SG: **Quantifying heterogeneity in a meta-analysis**. *Statistics in medicine* 2002, **21**(11):1539-1558.
70. Rücker G, Schwarzer G: **Ranking treatments in frequentist network meta-analysis works without resampling methods**. *BMC Medical Research Methodology* 2015, **15**(1):58.
71. Carlsen L, Bruggemann R: **Partial order methodology: a valuable tool in chemometrics**. *Journal of Chemometrics* 2014, **28**(4):226-234.
72. El-Basil S: **Partial Ordering of Properties: The Young Diagram Lattice and Related Chemical Systems**. In: *Partial Order in Environmental Sciences and Chemistry.* edn. Edited by Brüggemann R, Carlsen L. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006: 3-26.
73. Ciomek K: **Drawing Hasse Diagram**. In.: CRAN; 2017.
74. Tidermark J, Bergstrom G, Svensson O, Tornkvist H, Ponzer S: **Responsiveness of the EuroQol (EQ 5-D) and the SF-36 in elderly patients with displaced femoral neck fractures**. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2003, **12**(8):1069-1079.
75. Tidermark J, Ponzer S, Svensson O, Soderqvist A, Tornkvist H: **Internal fixation compared with total hip replacement for displaced femoral neck fractures in the elderly. A randomised, controlled trial**. *The Journal of bone and joint surgery British volume* 2003, **85**(3):380-388.
76. Blomfeldt R, Törnkvist H, Eriksson K, Söderqvist A, Ponzer S, Tidermark J: **A randomised controlled trial comparing bipolar hemiarthroplasty with total hip replacement for displaced intracapsular fractures of the femoral neck in elderly patients**. *Journal of Bone & Joint Surgery, British Volume* 2007, **89-B**(2):160-165.
77. Figved W, Opland V, Frihagen F, Jervidalo T, Madsen J, Nordsletten L: **Cemented versus uncemented hemiarthroplasty for displaced femoral neck fractures**. In: *Clin Orthop Rel Res.* vol. 467; 2009: 2426-2435.
78. Støen R, Lofthus C, Nordsletten L, Madsen J, Frihagen F: **Randomized trial of hemiarthroplasty versus internal fixation for femoral neck fractures: no differences at 6 years**. In: *Clin Orthop Rel Res.* vol. 472; 2014: 360-367.

79. Waaler Bjornelv GM, Frihagen F, Madsen JE, Nordsletten L, Aas E: **Hemiarthroplasty compared to internal fixation with percutaneous cannulated screws as treatment of displaced femoral neck fractures in the elderly: cost-utility analysis performed alongside a randomized, controlled trial**. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 2012, **23**(6):1711-1719.
80. Hedbeck C, Blomfeldt R, Lapidus G, Törnkvist H, Ponzer S, Tidermark J: **Unipolar hemiarthroplasty versus bipolar hemiarthroplasty in the most elderly patients with displaced femoral neck fractures: a randomised, controlled trial**. In: *Int Orthop.* vol. 35; 2011: 1703-1711.
81. Calder S, Anderson G, Harper W, Gregg P: **The use of bipolar hemi-arthroplasty in intracapsular hip fracture - a prospective randomised trial [abstract]**. In: *Journal of Bone and Joint Surgery - British Volume.* vol. 77 Suppl 1; 1995: 103.
82. Calder SJ, Anderson GH, Harper WM, Jagger C, Gregg PJ: **A subjective health indicator for follow-up. A randomised trial after treatment of displaced intracapsular hip fractures**. *J Bone Joint Surg Br* 1995, **77**(3):494-496.
83. Langslet E, Frihagen F, Opland V, Madsen J, Nordsletten L, Figved W: **Cemented versus uncemented hemiarthroplasty for displaced femoral neck fractures: 5-year followup of a randomized trial**. In: *Clin Orthop Rel Res.* vol. 472; 2014: 1291-1299.
84. Parker M, Khan R, Crawford J, Pryor G: **Hemiarthroplasty versus internal fixation for displaced intracapsular hip fractures in the elderly. A randomised trial of 455 patients**. In: *The Journal of bone and joint surgery British volume.* vol. 84; 2002: 1150-1155.
85. Skinner P, Riley D, Ellery J, Beaumont A, Coumine R, Shafighian B: **Displaced subcapital fractures of the femur: a prospective randomized comparison of internal fixation, hemiarthroplasty and total hip replacement**. *Injury* 1989, **20**(5):291-293.
86. Reuling E, Sierevelt IN, van den Bekerom MPJ, Hilverdink EF, Schnater JM, van Dijk CN, Goslings JC, Raaymakers E: **Predictors of functional outcome following femoral neck fractures treated with an arthroplasty: limitations of the Harris hip score**. *Arch Orthop Trauma Surg* 2012, **132**(2):249-256.
87. Tol MC, van den Bekerom MP, Sierevelt IN, Hilverdink EF, Raaymakers EL, Goslings JC: **Hemiarthroplasty or total hip arthroplasty for the treatment of a displaced intracapsular fracture in active elderly patients: 12-year follow-up of randomised trial**. *The bone & joint journal* 2017, **99-b**(2):250-254.
88. Frandsen P: **Osteosynthesis of displaced fractures of the femoral neck. A comparison between Smith-Petersen osteosynthesis and sliding-nail-plate osteosynthesis - a radiological study**. In: *Acta Orthop Scand.* vol. 50; 1979: 443-449.
89. Howie CR, Christie J, Watson H: **A PROSPECTIVE RANDOMIZED TRIAL OF COMPRESSION SCREW FIXATION VERSUS DOUBLE DIVERGENT PIN FIXATION FOR DISPLACED SUBCAPITAL FRACTURE OF THE FEMUR**. *Journal of Bone and Joint Surgery-British Volume* 1987, **69**(3):490-490.
90. Durlak JA: **How to Select, Calculate, and Interpret Effect Sizes**. *Journal of Pediatric Psychology* 2009, **34**(9):917-928.
91. Faraone SV: **Interpreting estimates of treatment effects: implications for managed care**. *P & T : a peer-reviewed journal for formulary management* 2008, **33**(12):700-711.
92. Macaulay W, Nellans K, Garvin K, Iorio R, Healy W, Teeny S, Yoon RSH, Geller J, Rosenwasser MP, Dfacto C: **Prospective randomized clinical trial comparing hemiarthroplasty to total hip**

arthroplasty: Functional outcomes in the treatment of displaced femoral neck fractures. *Osteoporosis Int* 2006, **17**:S238-S239.
93. El-Abed K, McGuinness A, Brunner J, Dallovedova P, O'Connor P, Kennedy J: **Comparison of outcomes following uncemented hemiarthroplasty and dynamic hip screw in the treatment of displaced subcapital hip fractures in patients aged greater than 70 years**. In: *Acta Orthop Belg.* vol. 71; 2005: 48-54.
94. Calder S, Anderson G, Harper W, Jagger C, Gregg P: **A subjective health indicator for follow-up. A randomised trial after treatment of displaced intracapsular hip fractures**. In: *The Journal of bone and joint surgery British volume.* vol. 77; 1995: 494-496.
95. Raia F, Chapman C, Herrera M, Schweppe M, Michelsen C, Rosenwasser M: **Unipolar or bipolar hemiarthroplasty for femoral neck fractures in the elderly?** In: *Clin Orthop Rel Res.* 2003: 259-265.
96. Blomfeldt R, Tornkvist H, Ponzer S, Soderqvist A, Tidermark J: **Internal fixation versus hemiarthroplasty for displaced fractures of the femoral neck in elderly patients with severe cognitive impairment**. *J Bone Joint Surg-Br Vol* 2005, **87B**(4):523-529.
97. Parker MJ, Blundell C: **Choice of implant for internal fixation of femoral neck fractures - Meta-analysis of 25 randomised trials including 4,925 patients**. *Acta Orthop Scand* 1998, **69**(2):138-143.
98. Parker MJ, Gurusamy KS, Azegami S: **Arthroplasties (with and without bone cement) for proximal femoral fractures in adults**. *Cochrane Database Syst Rev* 2010(6):CD001706.
99. Parker MJ, Gurusamy KS: **Internal fixation versus arthroplasty for intracapsular proximal femoral fractures in adults**. In: *Cochrane Database Syst Rev.* John Wiley & Sons, Ltd; 2006.
100. Katz JN, Phillips CB, Baron JA, Fossel AH, Mahomed NN, Barrett J, Lingard EA, Harris WH, Poss R, Lew RA *et al*: **Association of hospital and surgeon volume of total hip replacement with functional status and satisfaction three years following surgery**. 2003, **48**(2):560-568.
101. Ravi B, Croxford R, Hawker GA: **Low surgeon volume is associated with increased complications following total hip arthroplasty, after accounting for experience**. *Osteoarthritis Cartilage* 2013, **21**:S135-S136.
102. Zhang BF, Wang PF, Huang H, Cong YX, Wang H, Zhuang Y: **Interventions for treating displaced intracapsular femoral neck fractures in the elderly: a Bayesian network meta-analysis of randomized controlled trials**. *Sci Rep* 2017, **7**(1):13103.
103. Avery P, Baker R, Walton M, Rooker J, Squires B, Gargan M, Bannister G: **Total hip replacement and hemiarthroplasty in mobile, independent patients with a displaced intracapsular fracture of the femoral neck: a seven- to ten-year follow-up report of a prospective randomised controlled trial**. In: *The Journal of bone and joint surgery British volume.* vol. 93; 2011: 1045-1048.
104. Mosseri J, Trinquart L, Nizard R, Ravaud P: **Meta-Analysis of a Complex Network of Non-Pharmacological Interventions: The Example of Femoral Neck Fracture**. *PLoS One* 2016, **11**(1):16.
105. Liang CW, Yang FJ, Lin WL, Fan YQ: **Efficacies of surgical treatments based on Harris hip score in elderly patients with femoral neck fracture**. *Int J Clin Exp Med* 2015, **8**(5):6784-6793.
106. Abdelkhalek M, Abdelwahab M, Ali AM: **Bipolar versus fixed-head hip arthroplasty for femoral neck fractures in elderly patients**. *Strategies in trauma and limb reconstruction (Online)* 2011, **6**(1):1-6.
107. Davison J, Calder S, Anderson G, Ward G, Jagger C, Harper W: **Treatment for displaced intracapsular fracture of the proximal femur. A prospective, randomised trial in patients aged 65 to 79 years**. In: *Journal of bone & joint surgery - british volume.* vol. 83; 2001.

108. Koval KJ, Aharonoff GB, Rosenberg AD, Schmigelski C, Bernstein RL, Zuckerman JD: **Hip fracture in the elderly: the effect of anesthetic technique**. *Orthopedics* 1999, **22**(1):31-34.
109. Mattsson P, Larsson S: **Calcium phosphate cement for augmentation did not improve results after internal fixation of displaced femoral neck fractures: a randomized study of 118 patients**. *Acta Orthop* 2006, **77**(2):251-256.
110. Kumar S, Bharti A, Rawat A, Kumar V, Avasthi S: **Comparative study of fresh femoral neck fractures managed by multiple cancellous screws with and without fibular graft in young adults**. *Journal of clinical orthopaedics and trauma* 2015, **6**(1):6-11.
111. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT: **Evaluating the Quality of Evidence from a Network Meta-Analysis**. *PLoS One* 2014, **9**(7):e99682.