Predicting Outcome after total hip and knee replacement Study (POSt) – a prospective observational pilot study

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Mr Jeya Palan

BSc (Hons) MBBS FRCS (Tr&Orth)

College of Medicine, Biological Sciences & Psychology

University of Leicester

2020

Table of Contents

Dedication

To Rachael, Rubin and Olivia. With love.

Acknowledgements

I would like to express my thanks and gratitude to Prof Joseph Dias and Prof John Thompson, who were my supervisors and provided me with time, support and advice throughout this whole process and never lost faith!

I would also like to thank the following people, who are (past and present) members of the Academic Team of Musculoskeletal Surgery (ATOMS) based at Leicester General Hospital, without whom POSt (and this thesis) would not have been possible: Elaine James, Rosie Billany, Sabile Tahiri, Souzana Konstantinidou, Nicola Woofenden, Chris Bunce, Priya Radia, Mr Colin Esler and Prof Paul Watson. I would like to thank the British Medical Association (Doris Hillier grant) and the Foxtrot charity for helping to fund POSt.

At the very heart of POSt are our patients and this thesis is dedicated in part to all of the patients who have taken part in POSt. I would like to thank them for their time, support and enthusiasm in taking part in this study and without whom, there would be no research, no learning and no improvements. It is my sincere hope that the fruits of this research will lead to better outcomes for patients having hip or knee replacements in the future.

I would like to thank my parents for all of their support and encouragement over the years. Finally, I would like to thank my wife Rachael and my children Rubin and Olivia, who have supported me throughout the whole process and given me boundless love and affection and showed me that there was light at the end of the tunnel.

Thank you.

Glossary

ACL	Anterior cruciate ligament
AE	Adverse event
AR	Adverse reaction
BPI	Brief Pain Inventory
BMI	Body Mass Index
CCA	Complete case analysis
CF	Consent Form
CI	Chief Investigator
CRA	Clinical Research Associate
CRF	Case Report Form
DH	Department of Health
FJS-12	Forgotten Joint Score-12
GCP	Good Clinical Practice
GP	General Practitioner
HADS	Hospital Anxiety and Depression Scale
HHS	Harris Hip Score
HSS	Hospital for Special Surgery
KSS	Knee Society clinical rating system
KOOS	Knee Injury and Osteoarthritis Outcome Score
LOCF	Last observation carried forward
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random
MI	Multiple Imputation
MICE	Multiple Imputation by Chained Equations
NHS	National Health Service
OA	Osteoarthritis

OHS	Oxford Hip Score
∆OHS	Change in Oxford Hip Score
OKS	Oxford Knee Score
∆OKS	Change in Oxford Knee Score
PA	Physical Activity
PI	Principal Investigator
PIS	Participant/ Patient Information Sheet
PPI	Patient Public Involvement
PROMs	Patient Reported Outcome Measures
RA	Rheumatoid Arthritis
RCT	Randomised Control Trial
REC	Research Ethics Committee
SAPS	Self-Administered Patient Satisfaction score
SER	Self-Efficacy for Rehabilitation scale
SOP	Standard Operating Procedure
THR	Total Hip Replacement
TKR	Total Knee Replacement
TSK-11	Tampa Scale of Kinesiophobia-11
UK	United Kingdom
USA	United States of America

Predicting Outcome after total hip and knee replacement Study (POSt) – a prospective observational pilot study

Mr Jeya Palan

Abstract

Osteoarthritis causes joint pain over 8.5 million people in the UK and was the primary reason (>95%) for performing over 150,000 hip and knee replacements annually in the NHS. Up to 20% of patients continue to be dissatisfied following surgery. The reasons for this are multi-factorial and remain unclear. The Predicting Outcome Study after total hip and knee replacement (POSt) was designed to predict which patients are likely to have a poor outcome after surgery by assessing psychological factors, patient expectations and using accelerometers to provide an objective measurement of physical activity. The POSt Pilot study was conducted to help determine the general feasibility of the study, evaluate the study protocol, identify any problems with study conduct, assess the use of questionnaires and accelerometers and finally to determine the level of missing data and propose a method for mitigating and managing this in the data analysis plan.

The first 100 patients recruited to POSt were included in the pilot with 95 patients having data available for analysis at 6 months postoperatively. The pilot study had a recruitment rate of 30%, consent rate of over 70% and retention rates of 99%. An error with the use of one of the questionnaires (TSK-11) was identified and a plan of action explored to manage this error. Missing data in terms of body mass index (BMI), Oxford hip and knee scores was examined and an analysis plan for managing this issue proposed. Finally, the use of accelerometers in the study was assessed and found to be a valuable method of providing objective measurements of physical activity.

In summary, the results of the pilot study showed that the study protocol was appropriate, the study was feasible, the use of the questionnaires and accelerometers acceptable by patients and a missing data analysis plan was developed.

Chapter 1 Objectives of POSt-Pilot

1.1 Introduction

Osteoarthritis (OA) of the hip and knee affects millions of patients in the United Kingdom (UK) can be extremely painful and debilitating, adversely affecting a patient's quality of life(UK 2013a). Once non-operative measures such as analgesia, lifestyle changes (including weight loss and modification in activities), physiotherapy and joint injections, have been exhausted, the gold standard treatment for symptomatic OA remains joint replacement. Indeed, total hip and knee replacements are one of the most successful operations in modern surgery(Learmonth et al. 2007). Despite the success of total hip replacements (THR) and total knee replacements (TKRs), a significant minority of patients (10 - 20%) continue to be dissatisfied despite their surgery(Baker et al. 2007). The reasons for this are likely to be multifactorial with some factors that are potentially modifiable and others that are not(Robertsson et al. 2000; Scott et al. 2016; Williams et al. 2013; Anakwe, Jenkins & Moran 2011a; Bourne et al. 2009; Choi & Ra 2016; Bullens et al. 2001; Palazzo et al. 2014; Arden et al. 2011). A screening tool for clinicians would be helpful in predicting which patients preoperatively are likely to be dissatisfied after their surgery. If these risk factors can be identified, then specific interventions aimed at addressing them could be introduced before such a patient has their THR or TKR, thereby improving their outcome and satisfaction rates.

The Predicting Outcome Study after total hip and knee replacement (POSt) was designed in order to predict which patients were likely to be dissatisfied after THR or TKR. The role of the pilot study (POSt Pilot) was an essential part of the design of the main study (POSt Main) and forms the basis of this PhD thesis. As the study would involve the use of multiple questionnaires and accelerometers at different time points before and after surgery, the potential burden on patients needed to be explored and evaluated. Furthermore, POSt

Pilot would help provide valuable information on the general feasibility of POSt Main, the recruitment and retention rate of participants, identify any methodological flaws in the protocol and any deviation from the protocol, provide a method for dealing with missing or inaccurate research data and finally, provide recommendations for a data analysis plan for POSt Main.

1.2 Objectives

The objectives of POSt-Pilot were as follows:

1. To use a patient focus group as part of the process of Patient Public Involvement (PPI) in order to identify the importance of the research question, determine the potential questionnaire burden on study participants, explore the feasibility of using accelerometers and discuss potential areas for intervention in patients at higher risk of having a poor outcome.
2. To determine the general feasibility of the study in terms of the study protocol, patient recruitment, retention and attrition rates and to identify any weaknesses or errors in the study design and conduct.
3. To explore missing and inaccurate data and identify statistical processes to help mitigate the effect of missing data.
4. To create a data analysis plan based on statistical modelling designed to handle missing data from the dataset.
5. To analyse the pilot data from accelerometers as an objective measure of activity levels in relation to patient reported outcome measures (PROMs) and patient satisfaction.

1.3 Patient and public involvement (PPI)

Patient and public involvement (PPI) is now a necessary and essential aspect of any clinical research and funding application. It provides invaluable information for researchers in terms of optimising the design and conduct of any clinical study, putting patients at the very heart of the research process(Gillon 1994; Thornton 2008). The role of PPI in POSt Pilot was integral

to helping shape and modify the research protocol in order to determine the importance of the research hypothesis, optimise patient recruitment, minimise the potential burden of questionnaires and accelerometers and explore potential avenues for interventions based on the results of POSt. The use of PPI in helping with the study protocol was also an integral part in the successful grant application to the British Medical Association (BMA) leading to the award of substantial funding for the study (Doris Hillier grant). Chapter 3 of this thesis provides further information on the process of PPI and the importance of the patient focus group in helping to improve the research methodology of POSt Main.

1.4 General feasibility

One of the central aims of the POSt Pilot study was to review the study protocol and assess patient screening and consent, recruitment and retention rates, explore the potential burden of questionnaires and accelerometers, adherence to the protocol and the process of follow up and finally to identify any problems with the study protocol, design and conduct. The feasibility of the main study will be explored further and discussed in Chapter 4. In this chapter, the overall design and methodology of POSt Pilot is discussed in detail and the results of the feasibility analysis of the pilot is reported. One of the issues raised by the results of the pilot study was the use of an incorrect questionnaire, in this case, the Tampa Scale for Kinesiophobia-11 (TSK-11) which forms the basis of Chapter 5 and how such an error could be managed.

1.5 Inaccurate data

Inaccurate data can be a significant issue with any study, leading to loss of data, a reduction in the sample size and power of a study, thus rendering the results inconclusive and potentially subject to bias. The use of an incorrect questionnaire can lead to significant loss of data as the data collected would potentially be invalid, placing the whole study at risk of not being able to answer the research question posed, by reducing the sample size and power of the study. Chapter 5 discusses how an error in the TSK-11 questionnaire was

identified from the pilot study and a contingency plan put into place to see if data could be salvaged despite the error.

1.6 Missing data

Missing data is a common occurrence in any clinical research and optimising the study design and conduct is the best way to help reduce the burden of missing data. The handling of missing and inaccurate data requires a robust analysis plan exploring different statistical methods for dealing with missing data. Many questionnaires have rules governing how they should be scored and utilised and in particular, what to do when some of the data is missing. For example, the OHS and OKS allows up to two items to be missing but if more than two items have missing data, then the score cannot be calculated. Chapter 6 looks at using multiple imputation (MI) in order to mitigate against the effect of missing data and specifically explores the use of multiple imputation in association with data such as body mass index, where data is missing at a unit level. The chapter also explores how multiple imputation can be used to manage missing data at an item level in the case of the OHS and OKS. The chapter will make a series of recommendation to optimise the study design and conduct of POSt Main, in order to reduce the level of missing data being collected.

1.7 Accelerometers

Physical activity (PA) is the body movement produced as a result of the expenditure of energy by skeletal muscle(Caspersen et al. 1985) and measuring PA can be complex, involving the use of both subjective and objective measures. One of the more unique aspects of this study is its use of accelerometers in order to provide objective measurements of PA. Accelerometers are devices that measure the acceleration of an object in motion along lines of axes(Yang & Hsu 2010a). Chapter 7 of this thesis will set out to explore and analyze the preliminary data from the accelerometers in order to determine the feasibility of accelerometer use in terms of:

- reviewing the literature that currently exists in the use of accelerometer in patients who are undergoing hip and knee replacements.
- reviewing the practical application of using the accelerometers, recording any losses of the accelerometers, complications related to wearing the devices and how well patients tolerate wearing them.
- identifying one or two specific outputs from accelerometer data that could be used as markers of PA in any subsequent analysis.
- assessing if PA changes after TKR or THR using an accelerometer and if this data was also captured using the OKS/OHS, thus comparing subjective with objective PA data.

Chapter 7 will outline a series of recommendations as to whether or not accelerometers should be used based on the results from this pilot study and if so, how best to employ the accelerometers and which accelerometer outputs should be used in any subsequent data analysis of the larger POSt Main study. Due to the small sample size in the pilot study, it would not be possible to determine if there is a correlation between accelerometer data and questionnaire data. This would be possible with the main POSt Main study with an estimated sample size of approximately 400 patients, in particular exploring activity-based questionnaires (subjective data) with accelerometer data (objective data).

1.8 Conclusion

POSt Pilot will help identify the feasibility of the main study (POSt Main), assess methods of dealing with missing data from questionnaires using multiple imputation modelling, address the TSK-11 error, review the initial accelerometer data and provide a detailed statistical analysis plan for future work.

Chapter 2 Background

2.1 Epidemiology of osteoarthritis

Osteoarthritis (OA) is the most common form of arthritis in the world(Litwic et al. 2013). The reported prevalence of OA varies depending on which joint is affected and how OA is defined, both in terms of clinical presentation and radiological features. In the United Kingdom (UK) it is estimated that approximately 8.5 million people have joint pain due to OA(ARUK 2014). In England, approximately one in five adults aged 45 years and over have OA of the knee and the figure is around one in nine for hip OA(ARUK 2014). In the United States of America (USA), almost 27 million people had clinically diagnosed OA in 2005 and this figure is likely to have risen since then(R. C. Lawrence et al. 2008). The most commonly affected joints include the hand, knee and hip joints but OA can affect any joint in the body. The prevalence of osteoarthritis of the knee and hip can be variable. The Framingham Osteoarthritis Study in the US found a prevalence rate of 33% for radiographic OA (with a cohort of 1420 patients) and in the over 80's, the rate was 43% (Felson et al. 1997, Allen and Golightly 2015). More specifically, the same study found a prevalence of 19% for radiographic hip OA and 4.2% for symptomatic hip OA (Kim et al. 2014, Felson et al.1997). In a systematic review of 23 studies, Dagenais et al. found a wide range for the prevalence of radiographic hip OA, from 0.9% to 27% with a mean prevalence of 8% and a standard deviation of 7%(Dagenais et al. 2008). Studies from Bristol(McAlindon et al. 1992) and Nottingham{O'Reilly et al. 1996} have shown a prevalence of radiographic OA associated with knee pain, of 13% and 19% respectively. The impact of OA is considerable, affecting not only patients' quality of life but also from a health economics perspective, accounting for up to 2.5% of the gross national product of Western countries(Reginster 2002). Even though the prevalence of OA in the general population is already high, the burden of OA will continue to increase with an ageing population and obesity epidemic.

2.2 Definition of Osteoarthritis

Osteoarthritis (OA) is a chronic, degenerative disease of synovial joints, primarily affecting articular cartilage but also involving the surrounding structures that make up the synovial joint, such as bone, muscle, ligament and synovium (Litwic et al. 2013). Osteoarthritis can be defined either on the basis of radiological features or clinical presentation which includes joint pain and swelling associated with stiffness, a limited range of motion in the joint and crepitus(Hunter & Felson 2006). In 1957, Kellgren and Lawrence described the classic radiographic features of OA: loss of joint space, the presence of osteophytes, subchondral sclerosis and cysts(Kellgren & Lawrence 1963). The relationship between symptomatic OA and radiographic appearances of OA is unclear. Some patients experience severe pain despite very minimal radiographic changes whilst others may have severe radiographic OA features but remain asymptomatic(Neogi et al. 2009; Bedson & Croft 2008). This is likely due, in part, to the very subjective experience of pain with both physical and psychological components.

2.3 Risk factors for the development of OA

The aetiology of OA is multi-factorial and complex. Each of the major risk factors for OA will be discussed in turn below. It is important to note that OA is not a generic condition but rather, a final end point initiated by any number of triggers and that OA affecting one particular joint may not share the same risk factors as OA in other joints. Risk factors for OA of the hip and knee (Table 1) can be classified as being intrinsic (or person level) or extrinsic (joint related) factors(Allen & Golightly 2015).

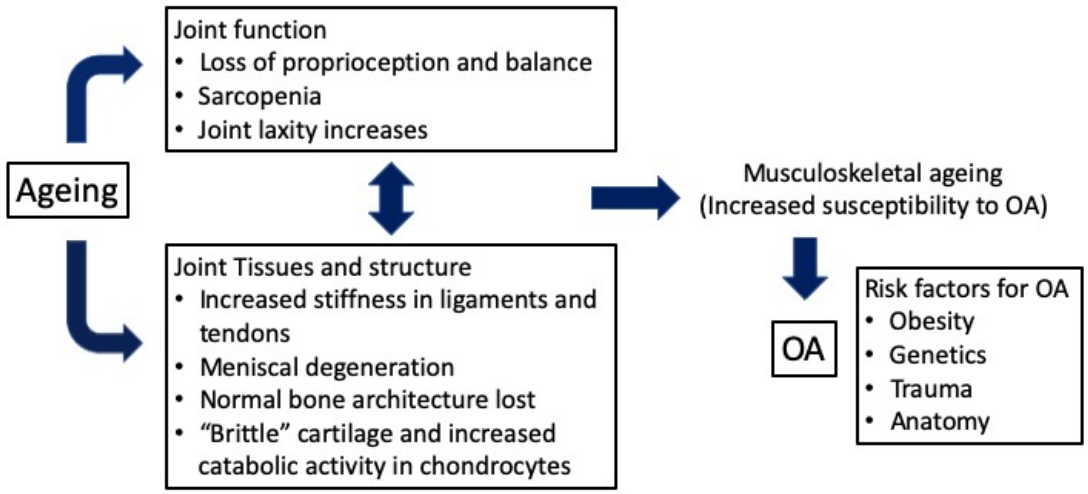
Table 1: Risk factors for OA (Allen & Golightly 2015)

Person specific	**Joint specific**
Age	Bone/Joint morphology
Gender	Trauma
Race and ethnicity	Muscle strength and mass
Genetics/Family history	Joint loads and kinematics
Obesity	Occupation
Nutrition and vitamin levels	Physical activity
Metabolic syndrome	Leg length inequality
Bone density	
Smoking	
Alcohol	
Cardiac disease	

2.3.1 Age

Increasing age is a strong predictor of the development and progression of OA but the mechanisms for this are not known. Age remains the strongest predictor of OA(Neogi et al. 2013) but increasing age does not necessarily lead to OA in all patients(Anderson et al. 2010). It is not the inevitable consequence of ageing. Age does affect the joint tissues and joint function that makes the joint more susceptible to developing OA. Sarcopenia, loss of proprioception and increased joint laxity all play a role(Johnson et al. 2014). The joint's ability to withstand articular injury and repair any damage becomes more limited with age.

Figure 1: The effects of ageing on the joint and the risk of developing osteoarthritis (OA) (Anderson 2010)



2.3.2 Gender

Women have a higher incidence of OA in knee compared to men (25-27). This incidence rate increases significantly in women after the menopause, leading to some researchers to postulate that knee OA was linked to sex hormonal levels such as oestrogen(Cirillo et al. 2006). This relationship remains unclear and unproven however(de Klerk et al. 2009). A meta-analysis by Srikanth et al. looking at 34 population based studies, found that the radiological severity of the knee OA was greater in women and that the impact of gender was greater in those aged 55 years and over(Srikanth et al. 2005).

2.3.3 Ethnicity

The incidence and prevalence of OA in the hip and knee appears to be influenced by race and ethnicity(Inoue et al. 2000, Dillon et al. 2006) Furthermore, other risk factors for OA may also vary depending on patient ethnicity. For example, the influence of a higher BMI is greater in Afro-Caribbean women compared to Caucasians(Wluka 2009). There are also racial and ethnic disparities in rates of THR and TKR with certain racial groups undergoing significantly lower rates of knee replacements compared to other

races(Skinner et al. 2003). This is most likely due to barriers in accessing healthcare linked with socioeconomic deprivation that are more prevalent in certain racial and ethnic groups within a system such as the USA, without a comprehensive federal healthcare model(Skinner et al. 2003).

2.3.4 Genetics

There appears to be a strong genetic component to the development and susceptibility of individuals to developing the disease(Lanyon et al. 2000). Family and twin studies have helped provide some information on the extent and influence of environmental factors and genetic factors on the development of OA(MacGregor & Spector 1999; MacGregor et al. 2008; MacGregor et al. 2000). Old age, ethnicity and female gender are also linked with OA(Blagojevic et al. 2010; Juhakoski et al. 2008).

2.3.5 Obesity

Obesity is established as one of the strongest modifiable risk factors for OA(Litwic et al. 2013) with greater BMI consistently being associated with increased risk of knee and hip OA(Blagojevic et al. 2010). There are several epidemiological studies that have shown a correlation between obesity, defined as having a Body Mass Index >30 and radiographic knee OA(Felson et al. 1997; Spector et al. 1994; Hochberg et al. 1995; Reijman et al. 2007; Cooper et al. 2000). The association between obesity (BMI > 30) and knee OA (Odds ratio OR 2.8; 95% CI 1.3-6.0) is significantly stronger than that with hip OA (OR 1.1; 95%CI 0.4-3.0)(Grotle et al. 2008). The knee joint is particularly vulnerable due to its mobility and because it's a weight bearing joint. It's stability depends on the strength of the ligaments and muscles which support it(Tortora & Derrickson 2010). The evidence for a link between obesity and hip OA remains contentious. Whilst some studies have shown a slight association(Cooper et al. 1998; Tepper & Hochberg 1993; Lievense 2002), others have shown little or no association at all(Grotle et al. 2008; Reijman et al. 2007). A significant proportion of patients undergoing a THR or TKR are obese and this may increase the risk of complications during and after joint replacement surgery,

including surgical site infections, greater blood loss, mal-positioning and nerve injury(Vasarhelyi & MacDonald 2012). The systematic review and meta-analysis by Blagojevic et al. (Blagojevic et al. 2010) on the risk factors for knee OA identified 36 papers that reviewed body mass index (BMI) as a risk factor and concluded that obesity was a significant risk factor for the development of symptomatic knee OA in the adult population with a pooled odds ratio (OR) of 2.63 (95% CI 2.28 – 3.05). Losing weight helped improve the symptoms of knee OA in obese patients(Felson et al. 1992). The reasons why obesity is associated with OA remain unclear. Mechanical factors and excessive joint loading play a role and there is some evidence that dyslipidaemia, low grade adipose tissue inflammation and the presence of adipokines are also implicated in obesity related OA(Thijssen et al. 2014).

2.3.5.1 Physical activity and OA

The role of physical activity and the risk of developing OA are less clear. The influence of sports activities and OA is uncertain although a systematic review by Lievense et al. suggested that there was probably a moderate association between sports activity and hip OA(Lievense et al. 2003). Certain manual occupations involving lifting, prolonged standing, kneeling or going up and down stairs may also be implicated in knee OA although the evidence is limited and based only on case control studies(Sutton et al. 2001; Coggon et al. 2000; Cooper et al. 1996; Manninen et al. 2002).

2.3.6 Post-traumatic OA

Trauma to the hip or knee joint has been associated with the development of OA(Wilder et al. 2002). In the knee, intra-articular fractures, anterior cruciate ligament (ACL) injury and meniscal injury can all lead to early onset OA. It is thought that the inflammation, mediated by pro-inflammatory mediators, that occurs directly after trauma, leads to chronic post-traumatic OA(Punzi et al. 2016). Post-traumatic OA may account for up to 12% of all OA cases(Punzi et al. 2016).

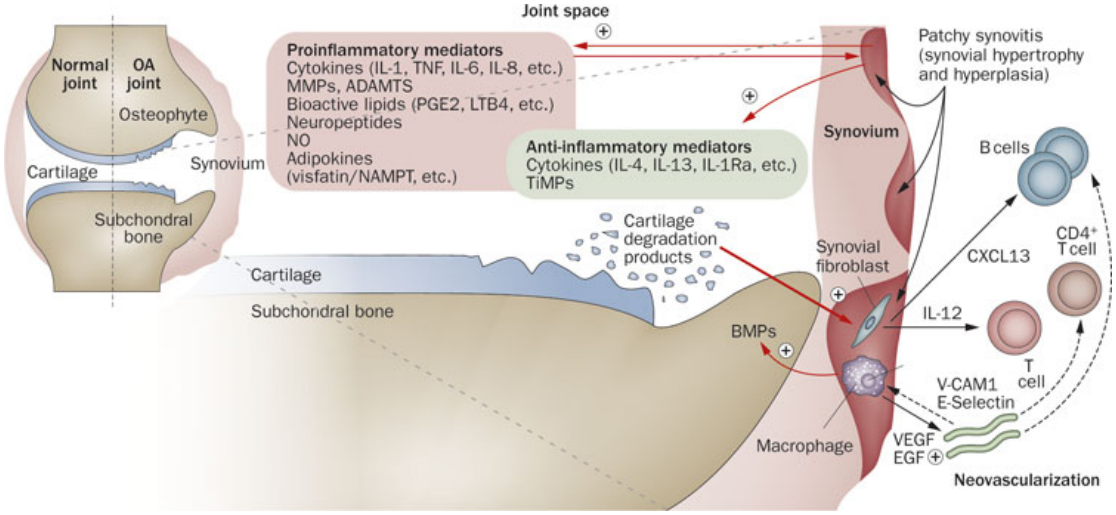
2.3.7 Mental health

The influence of mental health disorders and the development of knee OA are unclear although two studies suggest a slightly higher risk of knee OA in patients with depression or anxiety(Seavey et al. 2003; Palmer et al. 2007). The prevalence of clinically significant anxiety and depression in patient with OA has been estimated to be around 40% (around 2.5 times higher than the prevalence in the general population)(Axford et al. 2010). It is clear that psychological factors are strongly associated with pain, and thus symptomatic knee and hip OA.

2.4 Pathophysiology of OA

Although traditionally classified as a non-inflammatory arthritis (as compared to rheumatoid arthritis), it is now known that a significant synovial inflammatory component is often present, especially in more advanced cases of OA(de Lange-Brokaar et al. 2012). The reasons for this synovial inflammation remain contentious. The most widely accepted hypothesis is that fragments of cartilage from the degraded articular surface of the affected joint, come into contact with the synovium, triggering an inflammatory response as synovial cells are regarded as foreign bodies within the joint space. This pro-inflammatory response involves the release of synovial fluid mediators that in turn activate chondrocytes, leading to metalloproteinase synthesis and eventually further cartilage degradation(Mathiessen & Conaghan 2017) (Figure 2). Synovial angiogenesis is triggered by these pro-inflammatory mediators, which lead to the production of more mediators thereby setting up a vicious cycle of more and more cartilage degradation(Berenbaum 2013). Another theory is that the synovium itself is the primary trigger point for inflammation with synovial macrophages playing a key role in this process.

Figure 2: The role of synovitis in the pathophysiology of osteoarthritis (Mathiessen & Conaghan 2017) (Open Access permission)



2.5 Clinical and radiological features of OA

The diagnosis of OA is made using a combination of clinical and radiological criteria. The main symptoms of OA are usually pain and loss of function, primarily due to stiffness and swelling of the affected joint. Often, the symptoms are insidious and chronic in nature with patients only presenting to their general practitioner (GP) after months or even years of pain and disability. The presence of night pain, worsening function and failure of analgesia and other conservative measures are all indications that surgical intervention might be required. Radiographs of the affected joint typically show loss of joint space, the presence of osteophytes, subchondral bone cysts and sclerosis around the joint (See Figure 3). The Kellgren and Lawrence classification(Kellgren & Lawrence 1963) is the most commonly used system (Table 2) for describing the radiographic features of OA.

Figure 3: Radiographs (Antero-posterior) showing features of osteoarthritis in the (a) hip and (b) knee

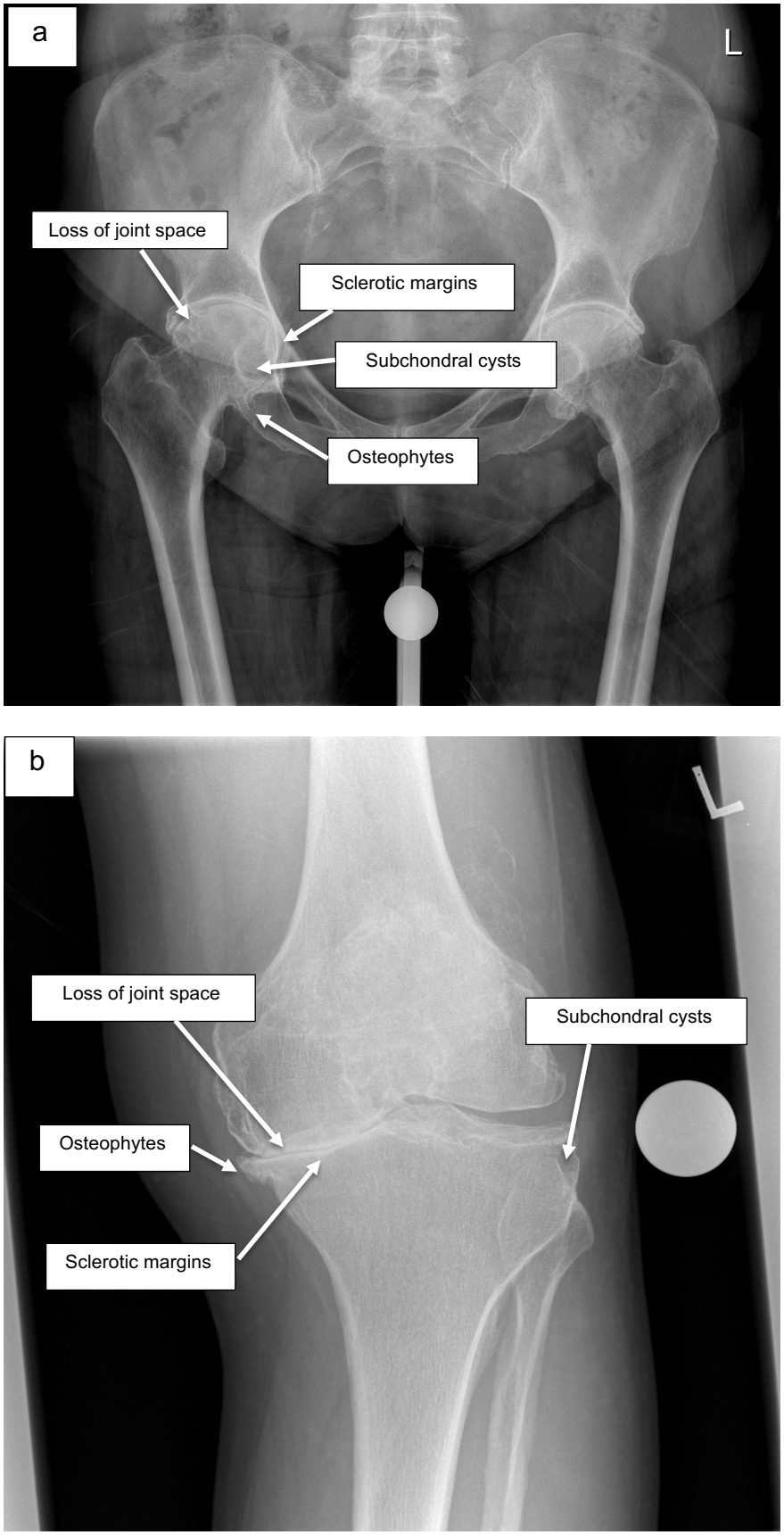


Table 2: Kellgren and Lawrence classification for radiographic appearances of osteoarthritis (OA) (Kellgren & J. Lawrence 1963)

Grade	Description
0	No radiographic features of OA seen
1	Possible joint space narrowing (JSN) and osteophytic lipping
2	Definite osteophytes and possible JSN on anteroposterior weight-bearing radiograph
3	Multiple osteophytes, definite JSN, sclerosis, possible bony deformity
4	Large osteophytes, marked JSN, severe sclerosis and definite bony deformity

For most patients, the presence of clinical symptoms and radiological evidence of OA is sufficient for the diagnosis of OA to be made. Occasionally, cross sectional imaging using CT or MRI scans may be required; if for example, there is abnormal anatomy or uncertainty in the diagnosis of OA. It should be noted that there is often a poor correlation between the radiographic appearances and clinical manifestation of OA.

2.6 Treatment of OA

Treatment for OA can be divided into two broad categories; non-operative or operative management. The aim of treatment is pain relief and an improvement in the function of that joint. In the first instance, it is worthwhile trying a course of non-operative treatment using a stepwise analgesic ladder process to control pain, together with physiotherapy, changes in lifestyle (including weight loss and stopping certain physical and sporting activities), using walking aids and joint injections. The presence of night pain sufficient to wake the patient up from sleep, unremitting pain despite strong analgesia, a significant loss of function with adverse impact on the patient's quality of life and ability to work and a failure of non-operative treatment to control the symptoms of OA, are all indications for surgical intervention.

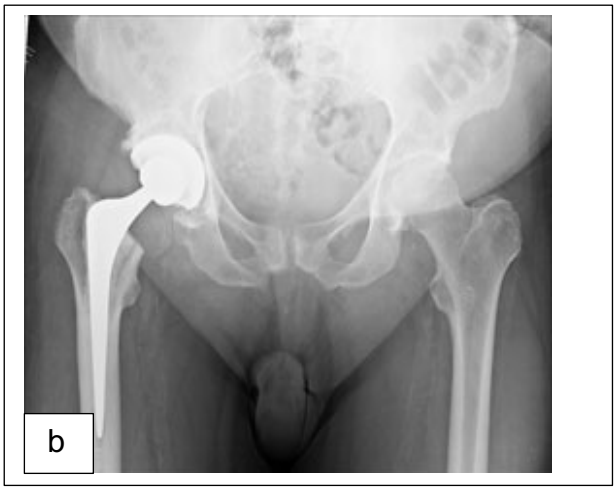
The role of arthroscopic debridement of the knee joint for OA of the knee remains controversial(Aaron 2006; Krych et al. 2014). The Finnish randomised controlled trial comparing sham surgery with knee arthroscopic debridement found the results of surgery to be no better than placebo(Kirkley et al. 2008) and its role in the surgical management of OA of the knee has diminished as a result. At present, biological treatments in the form of stem cell therapy and cartilage regeneration procedures remain an aspiration rather than a reality. The gold standard remains joint replacement surgery in the form of a total hip or knee replacement. Joint preservation surgery performing osteotomies around the knee or hip joint may be indicated in a select group of patients, who are usually young, very active and have a job that requires them to kneel or undertake heavy manual work. Fusion of the hip or knee joint or even excision arthroplasty of the joint is rarely indicated with the advent of joint replacement surgery. A total hip or knee replacement provides consistent, reproducible, successful outcomes in terms of pain relief and an improvement in function for patients with debilitating OA.

2.7 History and development of hip and knee replacements

Hip replacements have been around for over 100 years and have revolutionised the way severe painful osteoarthritis of the hip can be treated. The role of hip excision arthroplasty and fusion of the hip have been largely consigned to the history books because of the overwhelming success of hip replacements in transforming the quality of life of patients who were previously left crippled with pain and disability. The use of the first hip replacements is attributed to Glück, who in 1891 presented his series of ivory femoral head replacements in patients with tuberculosis of the hip (Learmonth et al. 2007). In the US, Smith-Peterson tried using glass (which unfortunately shattered under physiological loading) and then moved to stainless steel in 1925 in order to create the first total hip replacement (Smith-Petersen 1948). In 1953, an English surgeon, George McKee, was one of the first surgeons to use a metal on metal bearing for his total hip replacement. Then in the 1960’s, Sir John Charnley at Wrightington

(Manchester), introduced his concept of low friction arthroplasty, using a small metal femoral head on a metal stem with a polyethylene socket and using polymethylmethacrylate (PMMA) as bone cement(Knight et al. 2011). This continues to be the foundation of modern total hip replacements (Figure 4) to this day and the Charnley THR (in its various incarnations) continues to be widely used with extremely good success rates in terms of low revision rates at up to 25 years follow up(Allami et al. 2006). The types of bearings used in modern day THRs includes metal on polyethylene, ceramic on ceramic, ceramic on polyethylene and metal on metal, although this type of bearing has fallen into disrepute with the issues of metal debris particles and formation of “pseudotumours” and extensive soft tissue destruction around the hip joint as a result(Pandit et al. 2008). THRs can be cemented (in which both the femoral and acetabular components are cemented), uncemented (where both components are uncemented), hybrids (where the femoral component is cemented and the acetabular component is uncemented) and reverse hybrid (in which the acetabular component is cemented and the femoral component is uncemented).

Figure 4: Example of a total hip replacement (THR) (a) Exeter THR (Stryker) and (b) Radiographs anteroposterior view of the pelvis showing a right THR

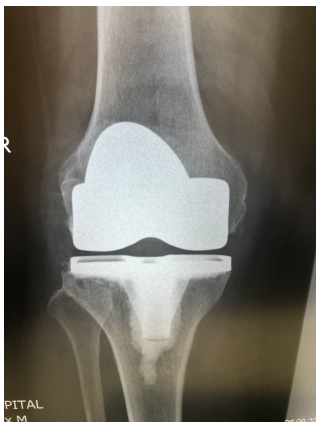


Total knee replacements (TKRs) have been around since the 1950's with Waldius using a simple hinge knee construct which had a high failure rate as it failed to take into account that the knee joint is a complex entity with movement both in flexion and extension, rotation and anterior-posterior glide{Song et al. 2013}. In 1971 Gunston developed a polycentric knee implant and other surgeons expanded on this design to produce a bicondylar knee replacement that matched the shape of the femoral condyles and tibial plateau more closely(Gunston 1971). Modern day knee replacements can be either cruciate retaining (CR) in which the posterior cruciate ligament (PCL) is preserved or posterior stabilised (PS) in which the PCL is sacrificed. Knee replacements can also be either mobile bearing or fixed bearing and can be implanted using cement or without cement (Figure 5). The survivorship of TKRs has improved over the last few decades and patients can realistically expect a TKR to last at least 15 to 20 years, and at least 80% will last more than 25 years(Evans et al. 2019).

Figure 5: Example of a total knee replacement (TKR) (a) Persona TKR (Zimmer Biomet) and (b) Radiographs Anteroposterior view of the Persona TKR

a

b



2.8 Measuring the outcome of joint replacement surgery

Primum non nocere (first, do no harm)

As with any medical intervention, there remains the possibility that such an intervention will be unsuccessful, will only be partially successful or in the very worst cases, cause the patient more harm than good. Measuring the outcome of any medical intervention is crucial in evaluating both the effectiveness and safety of the procedure and is a cornerstone of medical research, applicable to both the pharmaceutical and implant industry. Identifying what those outcome measures should be can be difficult and will vary depending on the type of intervention being undertaken. Furthermore, as medical interventions become more advanced and sophisticated, the outcome measures being measured may need to change over time. In the case of cancer therapy, the traditional biomedical outcome measures most relevant to the patient and clinician were patient survivorship or cancer cure rates. However, as cancer treatment has become more successful over time and survivorship has improved, such biomedical outcome measures, whilst still a fundamental measure of the success of cancer treatment, has expanded to include patient outcome measures including health related quality of life and the burden of treatment on patients.

In joint replacement surgery, the traditional model for evaluating success has been surgeon derived, primarily reviewing technical outcomes in terms of radiological parameters, survivorship of implants and presence of complications. Increasingly, the use of patient related outcome measures (PROMS) including patient satisfaction and validated patient-based outcome questionnaires has become the norm and patient satisfaction is now regarded as being a critical part of judging the success of surgery. It has been shown that there is often a discrepancy between surgeon and patient derived markers for success(Bullens et al. 2001). Any outcome measuring tool should be valid, reproducible and responsive to changes in the patient's condition(Bourne 2008). Outcome measures such as the Oxford Hip (OHS) and Knee Scores

(OKS) are disease specific, where as the EQ5D provides a more global assessment of health quality. The World Health Organisation Quality of Life study group recommended that any general health survey assess five domains; physical health, psychological health, social relationship attitudes, wellbeing and function. In our study, we have chosen to use the EQ5D, a well validated and reliable general health survey.

Some of these disease/joint specific scores are patient reported outcome measures (PROMs) and others are clinician reported. Some scores are applicable to both hips and knees whereas others are specifically designed for either the hip or knee. In the UK, the Oxford hip (OHS) and knee scores (OKS) are the most commonly used PROMs. The Oxford Hip(Dawson et al. 1996) and Knee Scores were developed as a patient reported outcome measures (PROM) in which patients complete a 12-item questionnaire measuring pain and function pre and post operatively after THR and TKR respectively. The OHS and OKS questionnaires are validated, patient self-reported with 12 questions, each with five options scoring from 0 to 4, all related to pain and function(Dawson et al. 1996; Dawson et al. 1998). The worst score is 0 and the best score is 48. These scores are simple to use and only take a few minutes to complete. The OHS and OKS are now used routinely in the NHS as part of the government's PROMs programme.

Other examples of outcome scores for the knee include the Knee Society clinical rating system (KSS) and the Knee Injury and Osteoarthritis Outcome Score (KOOS). The KSS was developed in 1989 and is a widely used clinician-reported scoring system used especially in the US for patients with knee OA(Insall et al. 1989) The clinical aspect of the KSS covers pain, range of movement (ROM), alignment and stability and is called the "knee score". The functional part (Function Score) of the KSS assesses the patient's mobility (walking distance and stairs) and any use of walking aids(Giesinger et al. 2014). The score range is from 0 to 100 points for each part with higher scores reflecting less severe pain and better function. The KOOS was originally

developed in 1998 for use in ACL reconstruction patients and is knee joint specific but has subsequently been validated for use in measuring outcomes after TKR also(Roos et al. 2003). The KOOS is a 42-item patient reported questionnaire designed to gauge a patient's perspective on any difficulties they may have with activities because of their knee problem. A higher score indicates a better outcome.

Harris Hip score (HHS) is one of the most commonly used instruments for assessing outcomes post-operatively after THR(Harris 1969) and is valid and reliable (Söderman et al. 2001). It is a clinician reported outcome measure. Another scoring system that is being used more frequently, is the Forgotten Joint Score-12 (FJS-12)(Behrend et al. 2012). The FJS-12 is a more recent PROM which was developed in 2012 and designed to measure the ability of the patient to "forget" about their joint after THR and TKR during various activities of daily living. It is based on a 5-point Likert response format and consists of 12 equally weighted questions. Up to four missing values are regarded as acceptable when the scores are summarized and transformed to a scale ranging from 0 to 100. High scores indicate good outcome, i.e., a high degree of being able to forget about the affected joint in daily life. It has been shown to have a low ceiling effect and excellent internal consistency (with a Cronbach's Alpha 0.95) and showed a good ability to discriminate well between different patient groups with different outcomes(Behrend et al. 2012).

2.9 Patient satisfaction after THR and TKR

Although THR and TKR have been generally regarded as being highly successful procedures in terms of relieving pain and improving function, patient satisfaction still remains a concern, especially after TKR. Patient dissatisfaction after knee replacements has been shown to vary between 10 and 20%(Nam et al. 2014; Bourne et al. 2009; Baker et al. 2007; Bryan, Goldsmith, Davis, Hejazi, MacDonald, McAllister, Randall, Suryaprakash, Wu & Sawatzky 2018a). The registry based study by Baker et al found that up to 18% of patients were dissatisfied after the TKR(Baker et al. 2007).

In contrast, patient satisfaction after THR is higher at around 91% to 93%(Arden et al. 2011; Anakwe, Jenkins & Moran 2011a). In a large prospective cohort study of 850 patients having a THR, the dissatisfaction rate was 7%(Anakwe, Jenkins & Moran 2011b). The reasons for this difference in satisfaction rates between THR and TKRs remain unclear(Bachmeier et al. 2001) and the POSt study hopes to determine some of the reasons why. One hypothesis is that the hip joint is a deep structure with a greater soft tissue envelope compared to the knee joint. After a THR or TKR, swelling, pain and a haemarthrosis are to be expected but as the hip joint is deep, much of this is contained within the soft tissue envelope. The knee joint, in contrast, is a superficial structure and even small amounts of swelling can lead to severe pain and loss of joint movement, which in turn can lead to stiffness and loss of motion. Satisfaction after surgery is strongly correlated to pain after surgery and this may account for difference seen in patient satisfaction between a THR and TKR.

The reasons for patient dissatisfaction after THR and TKR are multifactorial and complex in nature. This includes patient expectation mismatch, psychological factors (including anxiety, depression, pain catastrophising, kinesiophobia and self-efficacy), physical factors such as other joint or back pain and multiple comorbidities and finally non-medical factors including car parking, overall hospital care and the quality of hospital food. The chronicity of the condition has also been shown to influence patient satisfaction, with patients with more chronic pain conditions such as rheumatoid arthritis (RA) having higher satisfaction after TKR(Bullens et al. 2001) compared to patients with OA or more acute presentations such as post traumatic arthritis or avascular necrosis.

2.9.1 Multiple joint pain and back pain

The presence of pain in other joints or back pain has been associated with poorer outcomes after hip and knee replacements. Localising pain to a particular joint can be difficult in the presence of multiple joint pain or back pain. Referred pain on the ipsilateral side following THR or TKR can cause residual pain in that leg even after surgery and this may affect patient satisfaction. It is

not surprising that patients with arthritis in one joint may also have arthritis affecting other joints including the spine, since the degenerative process is similar. Back pain has been found to be present in almost 50% of patients undergoing THR(Parvizi et al. 2010) and in 35% of patients having a TKR(Clement, D. MacDonald & Simpson 2013). The presence of pain in other joints or concomitant back pain has been shown to influence patient satisfaction after THR or TKR. The presence of back pain was found to be an independent variable leading to lower satisfaction rates after TKR(Clement, D. MacDonald & Simpson 2013; Escobar et al. 2007). The presence of symptomatic arthritis in other joints was an independent predictive factor for dissatisfaction after THR(Anakwe, Jenkins & Moran 2011a). Pain and function have been shown to be strongly predictive of dissatisfaction after TKR(Baker et al. 2007). One of the difficulties in assessing PROMS questionnaires is isolating the affected joint from symptoms arising from other joints. For example, patients often find it difficult to determine if their inability to kneel after a TKR is because of the TKR or because they also have significant arthritis affecting their contralateral knee. Similarly, a patient may find continue to get pain in their hip as a result of referred pain coming from the lumbar spine.

2.9.2 Comorbidities

The impact of other medical comorbidities on patient satisfaction after THR or TKR remains unclear. Physical comorbidities do not appear to predict poorer outcome after TKR although pre-existing mental health comorbidities may influence outcome in a negative manner(Ayers et al. 2005). Another, more recent study suggests, however, that physical health preoperatively, is a predictor of satisfaction after TKR at six months postoperatively, together with preoperative pain and mental health(Bryan, Goldsmith, Davis, Hejazi, MacDonald, McAllister, Randall, Suryaprakash, Wu & Sawatzky 2018b).

2.9.3 Expectations

A mismatch between the patient and surgeon’s expectations of the outcome of THR or TKR can lead to patient dissatisfaction(Moran et al. 2003). The primary

reason for performing a THR or TKR is to alleviate pain and improve function. A prospective study of 1703 patients undergoing a primary TKR found that the greatest predictors of dissatisfaction were unrealistic patient expectations, the presence of a postoperative complication requiring hospitalisation, high preoperative pain scores at rest and a poor postoperative knee functional outcome score(Bourne et al. 2009). The importance of realistic patient expectations on patient satisfaction after surgery is supported by a number of other studies(Nilsdotter et al. 2009; Scott et al. 2012; Lingard et al. 2006; Haanstra & van den Berg 2012; Mancuso, Salvati, Johanson, Peterson & Charlson 1997a). Patients with poorer pre-operative function before THR had higher expectations for their operation and this may lead to higher rates of dissatisfaction in this group of patients(Mancuso et al. 2003). Despite a number of studies investigating the effect of expectation on patient satisfaction after TKR, a systematic review by Barlow et al. concluded that despite all of these studies, none of the studies undertook the appropriate sub group analysis and therefore, it was difficult to ascertain if expectation mismatch contributed significantly to patient dissatisfaction(Barlow et al. 2016). The authors recommend that better high quality research is required in this area and that consensus is needed on how best to measure expectation and satisfaction.

2.9.4 Persistent pain

In a large registry based study by Baker et al., the strongest predictors of dissatisfaction in patients having a primary TKR was persistent pain after surgery and that poor pain and function knee scores were associated with higher levels of patient dissatisfaction(Baker et al. 2007). Since the primary reason for performing a joint replacement is the relief of pain and improvement in function, the presence of persistent postoperative pain and poor function is naturally frustrating and disappointing for the patient and would therefore affect their satisfaction following surgery. In a small registry-based study, 114 patients who were very satisfied after their TKR 8 to 13 years after their surgery, were compared with a matched group of 113 patients who were dissatisfied with their TKRs and were still dissatisfied even after many years(Ali et al. 2014). The only

difference found between the 2 groups was that the dissatisfied group had a higher pain score and higher levels of anxiety and depression.

2.9.5 Mental health and preoperative psychological factors

There is a growing body of evidence to suggest that a patient's psychological profile, such as anxiety and depression, may also affect patient satisfaction and outcomes after joint arthroplasty(Clement, D. MacDonald & Burnett 2013; Lingard et al. 2004; Wylde et al. 2007; Brander et al. 2007). Psychological predictors such as high pain catastrophizing(Forsythe et al. 2008; Witvrouw et al. 2009), high fear avoidance beliefs, kinesiophobia and low self efficacy(van den Akker-Scheek et al. 2007), have been highlighted as potentially being important in relation to poor outcome and all are associated with a failure to become physically active after surgery{Ayers:2004va}.

2.9.6 Anxiety and Depression

The prevalence of psychological distress or other forms of mental health disability such as depression or anxiety in patients with a diagnosis of osteoarthritis is well recognised. In the study by Axford and colleagues, a prevalence of 40% for significant depression or anxiety was observed in 54 patients who attended a rheumatology clinic with a diagnosis of osteoarthritis(Axford et al. 2010). In a registry based study by Ali et al., 42% of patients who were dissatisfied after TKR had clinically significant anxiety or depression related symptoms compared to only 10% of patients in the satisfied group(Ali, Sundberg, Robertsson, Dahlberg, Thorstensson, Redlund-Johnell, Kristiansson & Lindstrand 2014b). Blackburn et al. found in their prospective study of 40 patients undergoing TKR, that more severe levels of preoperative anxiety and depression was associated with worse knee disability and that an improvement in post-operative levels of anxiety and depression led to improvements in knee disability(J. Blackburn et al. 2012).

2.9.7 Self-efficacy

Self-efficacy can be defined as the “conviction that one can successfully execute the behaviour required to produce the desired outcome” (Bandura 1977). A perception of self-efficacy engenders a sense of control over events and the belief that one’s behaviour (increasing activity levels following surgery) will lead to a desired outcome: in this case improved mobility or reduced pain(Hartley et al. 2008). Regardless of when self-efficacy is measured it has been demonstrated to influence the report of disability at short and long term follow-up in subjects undergoing hip or knee arthroplasty(van den Akker-Scheek et al. 2007; Wylde et al. 2007). A Dutch prospective study examined the effect of pre- and post-operative self efficacy in patients having a THR or TKR and found that short term post-operative self-efficacy was the strongest predictor of long term functional outcome after surgery(van den Akker-Scheek et al. 2007).

2.9.8 Kinesiophobia

Kinesiophobia is defined as a pain related fear of movement and is associated with the development of disability in musculoskeletal disorders(Heuts, Vlaeyen, Roelofs, de Bie, Aretz, van Weel & van Schayck 2004a). People with high fear of activity are more likely to avoid pain-provoking activity, limit physical performance and not engage in challenging activities(Somers et al. 2009). Through this behaviour the patient may initially successfully avoid a painful experience but the cost is maintaining disability or increasing the risk of disability. High levels of kinesiophobia have been shown to be associated with chronic lower back pain and prolonged disability(Picavet 2002).

2.9.9 Pain catastrophizing

Pain catastrophizing is characterised by a tendency to magnify the threat value of pain and in turn develop a sense of helplessness in the face of pain. It has been shown to predict postsurgical pain and post arthroplasty pain in particular(Riddle et al. 2011). Initial pain severity and scores on the Pain Catastrophizing Scale (PCS) can reliably predict pain severity 6 weeks after

TKR and might be predictive of chronic pain, and functional status(Forsythe et al. 2008). Those scoring high on the PCS had higher disability. As pain catastrophizing is strongly related to pain report and disability, treatments such as cognitive behavioural therapy that aims to reduce catastrophic thoughts demonstrate reductions in pain and disability(Riddle et al. 2011).

2.9.10 Physical activity levels

Preoperative physical activity (PA) has been associated with outcome after total hip and knee replacements for patients with OA and some cohort studies have reported better outcomes after surgery if patients are more physically active preoperatively(Pozzobon et al. 2018, Pietschmann et al. 2013). Other studies have shown no relationship between preoperative PA and outcomes(Poortinga et al. 2014). These studies have all used self-reporting questionnaires to measure PA which is very subjective. POSt will assess the association between PA and predicting outcome after TKR or THR using accelerometers as an objective measure of PA.

2.10 Interventions to improve outcome after THR and TKR

There have been a number of interventions that have tried to improve patient satisfaction and outcome after THR and TKR, based on studies which have identified some of the potentially modifiable risk factors for patient dissatisfaction. In particular, interventions focusing on influencing patients' psychological behaviours such pain catastrophizing and kinesiophobia as well as targeted exercise regimes and physiotherapy to improve rehabilitation engagement and compliance and patient education programmes designed to temper unrealistic patient expectations, have been tried with limited success.

Psychological factors are now thought to play a significant role in determining outcome after THR and TKR. Pain is a major factor in patients being dissatisfied after surgery. Psychological traits such as self-efficacy, pain catastrophizing and kinesiophobia have been identified as being possible targets for

psychological interventions such as cognitive behavioural therapy (CBT), relaxation methods, guided imagery, psycho-education and motivational interviewing. Dixon et al. conducted a meta-analysis looking at psychological interventions for arthritis pain management(Dixon et al. 2007). There was an overall effect size of 0.177 (95% CI 0.256–0.094) suggesting that patients who received psychosocial interventions had significantly lower pain than patients in the control group ($p < 0.01$). The study also supported the use of psychosocial interventions for improving psychological, physical, and biological functioning function.

In a systematic review of psychological interventions in THR and TKR, Bay et al. screened almost 19500 studies and included 7 studies for the review(Bay et al., 2018). Of these, two studies found psychological interventions to be effective in improving at least one patient reported joint outcome. The interventions were CBT with relaxation therapy in one study which improved function after THR(Berge et al. 2004) and guided imagery in the other study, which reduced pain 3 weeks after TKR(Jacobson et al. 2016). Overall, however, the authors concluded that despite increasing evidence that psychological traits played a significant role in predicting poor outcome after THR and TKR, further RCTs with focused interventions were needed(Bay et al. 2018).

A systematic review of RCTs was undertaken by Ackerman and Bennell to evaluate the effectiveness of pre-operative physiotherapy programmes on outcome following THR and TKR(Ackerman and Bennell, 2004). They found that pre-operative physiotherapy intervention did not improve outcome after TKR but their effect on outcome after THR could not be determined. In another meta-analysis, Gill and McBurney found that exercise-based interventions can improve outcome (reduced pain and better function) in THR patients but not TKR patients(Gill and McBurney, 2013). A systematic review, conducted as part of an NIHR programme grant (Improving patients' experience and outcome of total joint replacement: the RESTORE programme), was performed to

evaluate the clinical effectiveness of exercise and education in patients waiting for THR and TKR(Blom et al. 2016). The study authors found that exercise and education before surgery can improve patients' pre-surgical health and improve early recovery post-operatively, although this effect was mainly seen in patients having a THR and the longer term effect on outcomes remains unclear.

Unmet patient expectations are a significant factor in patient dissatisfaction especially after TKR. Previous research has shown that a structured pre-operative education programme on realistic expectations for long-term recovery after THR and TKR can influence pre-operative expectations(Mancuso et al. 2008). The EKSPECT (the influence of Expectation modification in Knee arthroplasty on Satisfaction of Patients) study is an RCT assessing whether an educational module on long-term recovery after TKR will improve patient satisfaction compared to the standard pre-operative patient education(Tolk et al., 2018). This study was registered in 2018 and its results have yet to be published.

One reason why these interventions may not have shown as positive a benefit is because they are being applied to all patients having a THR or TKR and overall, the vast majority of these patient will have a good or excellent outcome and be very satisfied. As a result, interventions applied in a blanket fashion may not show as great an impact whereas if patients who have been identified as being at higher risk of a poorer outcome could be targeted for such interventions, then for these patients, their outcome may be significantly improved as a result.

2.11 Conclusion

By developing a screening tool to predict which patients are likely to have a poorer outcome than expected, in terms of patient satisfaction and Oxford hip and knee scores post operatively (where poor outcome has been defined as

being equal or less than 27, where 0 is the worst score and 48 is the best), potential interventions can be instigated to help target such patients for specific help. Recommendations are for simple low cost interventions to address these obstacles to recovery and improve outcome(Westby 2012). Blanket approaches to intervention where participants are included by health condition rather than by risk of poor outcome have been criticised in the management of other conditions(Dorr & Chao 2007). Inclusion of people with low risk of poor outcome inevitably reduces the effect of the treatment. Previous research into screening for risk of poor outcome has demonstrated that those with a high risk for poor outcome specifically respond well to treatments designed to address these factors(Riddle et al. 2011). By undertaking a pilot study, the feasibility of the main POSt study could be evaluated, the design and conduct of the study protocol reviewed, the impact of multiple questionnaires as a potential burden for patients assessed and the use of accelerometers trialed.

Chapter 3 Patient Public Involvement

3.1 Introduction

Patient public involvement (PPI) has been defined as “experimenting with” rather than “experimenting on” patients involved with clinical studies(Hanley et al. 2004). It places the patient at the heart of clinical research as ultimately, the whole purpose of such work is for the benefit of the patient. The medical ethicist, Raanan Gillon stated that there was a moral and scientific imperative that patients and clinicians come together to form a “brave new partnership”, with both sides understanding the objectives and interests of one another with a willingness to compromise in the event of any conflict(Gillon 1994). In one of the earliest examples of PPI in action in the United Kingdom, the Association for Maternity Services in the 1980s, brought together voluntary organisations and patients’ groups in order to gain their support for a randomised controlled trial on chorionic villus sampling in pregnancy, with members of these groups directly involved in conducting and promoting the study(Chalmers 1986). The importance of PPI cannot be overstated and certainly from a funding perspective, many research funding bodies now require that researchers provide evidence of PPI when submitting their proposals including detailed plans for the role of PPI in every element of the proposed research. Indeed, from an ethical perspective, patients lie at the very heart of any and all clinical research(Gillon 1994). There is a strong moral argument on the grounds that people affected by a condition have a right to be directly involved in decisions about research that may affect them(Domecq et al. 2014). There are several other reasons why PPI is critically important when designing and indeed running a clinical study(Blackburn et al. 2018; Hanley et al. 2004).

- Provides a platform upon which clinicians and patients come together to discuss and develop research ideas and for many studies, the research question was proposed by a patient which then subsequently led to the research study being conducted in order to answer that research question. PPI helps provide evidence that the research question is important and relevant from a patient’s perspective. A good example of

this is the James Lind Alliance, an organisation set up in 2004 with the purpose of bringing together clinical researchers, patients and carers in order to identify areas of clinical research priorities in a range of different medical specialties.

- Helps direct and focus a research study, guiding and shaping the research methods. PPI can contribute to the writing of the research study protocol.
- Provides a valuable insight into the language and design of patient facing documents such as the patient information sheet (PIS), consent forms, patient questionnaires and diaries.
- Helps ensure that the study is conducted in a participant friendly and ethically manner
- Identifies potential pitfalls in the use of study materials, equipment, questionnaires and other research tools, for example, in the use of plain English when designing patient facing documentation such as the patient information sheet (PIS).
- Provides a public perspective on interpreting the research study's findings
- Helps disseminate the study findings to a broad audience including trial participants, clinicians and the public
- Provides a measure of the impact of a study's findings potentially helping to shape and inform future trial design

In a review of the impact of PPI in primary care research, Blackburn et al. have summarised the costs and benefits framework of PPI engagement with reference to the researcher, research project, research institution, funder and from the PPI participant's perspective (Table 1)

Table 1: Costs and consequences framework(adapted from Blackburn et al. 2018)

Impact upon		Cost (-)	Benefits (+)
Researcher		• Time (recruiting PPI contributors; travelling to meet with PPI contributors; meetings; electronic communication; preparing newsletters) • Increased pressure/stress • Sensitivity to criticism	• A motivating factor, with PPI contributors bringing an enthusiasm to the project, a keenness to see results • PPI contributors supportive of the project • Researchers gaining a better understanding of the condition of interest
Research project	Shaping the research question and maintaining focus		• Setting and maintaining focus on the research question • Addressing important issues but also ensuring a degree of realism
	Research methods/design	• Can result in duplication of effort (PPI involvement and qualitative work)	• Helping to make surveys and processes relevant, accessible and acceptable • Ensuring research is beneficial to patient group
	Recruitment & recruitment materials	• Potentially homogenous sample	• Relevance, clarity & accessibility of recruitment materials • Making useful contacts, increasing recruitment rates
	Conducting & managing research	• PPI contributors can be unreliable (this was reported in the case of young people) • Direct payment of PPI contributors for attending meetings • Travel costs (either the researcher visiting	• Validity and safety of research • Improved follow-up rates

		• the PPI representative or the PPI representative attending meetings[a]) • Food and refreshment costs • External venues	
	Commenting on results		• Opportunities to gain feedback and to validate the results. • PPI contributors helping to interpret the data
	Dissemination	• *Financial cost of PPI contributors attending conferences and external events*	• Promotion of outputs when these take the form of training modules or tool kits • Guidance in terms of presenting results in a format useful to non-researchers
	Generating new research questions (expanding upon current research)		• Generating new/future research questions
Research Institution		• Diversion of research funds to PPI (opportunity cost in terms of funded researcher time, etc.) • IT and other support infrastructures/resources (including printing & internal room bookings)	• Increased impact of research • Recognition as a centre with expertise and experience of involving patients and public in research (raising the institution's profile)
Funder			• Avoiding devoting resources to a topic which is not important (e.g. exploring an intervention which is not appealing to service users)
PPI contributors		• Opportunity cost (paid work, child care, informal care & leisure time) • Monetary costs not reimbursed (travel, formal child care)	• Increased understanding & knowledge of one's own condition • Increased awareness of treatment options and how to access services

		• Negative impact on health associated with stress, anxiety or frustration • Complications in terms of state provided welfare payments	• Developing or enhancing skills (e.g. public speaking, team work, IT) – possibly through formal training • Understanding of research and research processes • Positive emotional impact associated with meeting new people, feeling as though one is doing something worthwhile and generally being active

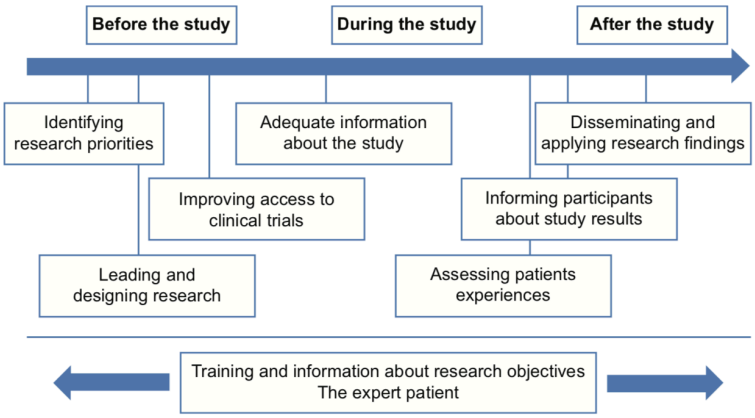
Entries in italics were identified from the literature but not verified by respondents

[a] Sometimes included within the direct payment

The role of PPI can take many forms, ranging from patient focus groups and patient surveys to having patients on research trial steering and management groups, being involved in community engagement and being authors on research papers and presentations and patient advisory groups(Blackburn et al. 2018).

There are three levels of PPI in research, as defined by INVOLVE. Consultation, collaboration and user led involvement with user led research, which is the highest and “best” level of involvement(Staley et al. 2012).

Figure 1: How to engage patients in clinical research(Sacristan et al. 2016)



Patient focus groups are group discussions organised in order to provide a patient’s perspective on the proposed research study and to help explore a number of different issues within the study(Kitzinger 1995). The term “focus” implies that the collective group concentrates on a particular topic or task such as reviewing research questionnaires or the research consent form(Kitzinger 1994). Focus groups have several advantages. They are an excellent method of collecting qualitative data based on the discussions between participants which help to generate ideas and recommendations(Kitzinger 1994). They can help provide high quality data as there is time to probe deeper into issues and

help develop ideas and themes about the research study(Wong et al. 2008). Finally, they can be hypothesis generating(Wong et al. 2008).

Focus groups also have limitations (Table 2) (Wong et al. 2008). They are subject to bias as by definition, the participants are being asked to provide their own personal perspective on a subject matter or issue. Furthermore, opinions may be swayed by dominant participants or indeed moderators, whose personal viewpoint may overshadow the views of other participants. It is therefore imperative that the moderators of the focus group ensure that all views are heard, especially from participants who are shy or quiet and less domineering. The moderator must also refrain from unduly influencing the discussion or opinions of the participants. Focus groups, if not carefully managed by the moderator, can also produce data which is unfocused as participants digress from the chosen issue. This can make data interpretation from focus groups more difficult.

Table 2: Focus group limitations and ways to mitigate such limitations(Wong et al. 2008)

Limitation	**Mitigation**
Participant bias Moderator bias	Moderator role in ensuring all views are heard Record all comments in full, from participants and moderators, and transcribe verbatim
Dominating participant(s)	Ensure homogenous group of participants Moderator involvement
Unfocused data Digression from core themes	Moderator role in managing discussion themes and topics
Difficult to organise and poor response rate	Clear venue and date and time Reminder letters and phone calls to participants beforehand Over recruit by 20% to allow for some participant drop out on the day of the focus group

3.2 Methods

In order to obtain a greater patient perspective into the importance of the research study question and the potential burden of questionnaires and accelerometers on patients, it was decided that a patient focus group would be set up. Recruitment took place from the outpatient clinic setting involving patients who were awaiting a total hip or knee replacement. Eight patients were selected at random, by reviewing the waiting lists for patients who were scheduled to have either a THR or TKR and invited to participate in the focus group. With funding secured from the NIHR PPI East Midlands Clinical Research Network (CRN) and the help of *MC* (Patient Advisor at UHL), the focus group was conducted on 1st June 2012. This involved *PW* (Research study group member) and *JP* (PI for POSt) chairing the focus group and five patients (as three patients withdrew from the group) as well as *MC* (patient liaison adviser for the hospital) as an independent patient observer. Our patient liaison adviser (MC) was also involved in helping develop the themes and content for the PPI focus group and this was based on a review of the literature on running a focus group and also as a result of my participation in an NIHR PPI educational day whereby I was able to receive advice on how to organise and run a PPI focus group as well as developing objectives for the focus group.

The focus group comprised of patients who have been listed for a primary THR or TKR (at the Glenfield or Leicester General Hospitals). Initially, I presented a short summary of the study as a PowerPoint presentation including the aims, objectives and study methodology, ensuring this was presented in layperson's terms. Patient feedback about this focus group was also gathered using a patient feedback survey, that was designed by myself and the wider . The proposed questionnaires for use in the study were available for the members of the patient focus group to review and make comments on. There were also accelerometers available for the patient members of the group to try on and review. Comments were recorded by the participants on feedback forms. *JP* explained in detail that this study consisted of two parts; the first, would help

identify the modifiable risk factors that would predispose certain patients to having a poorer outcome than expected. Once such modifiable risk factors had been identified, the next step would be to explore different treatment strategies to help mitigate these risk factors to help improve patient outcome and undertake clinical trials in order to test the success of these treatment options.

The PPI focus group involved collecting both targeted and open-ended responses from patients. The targeted responses were primarily applied to asking the patients to comment specifically on the questionnaires and accelerometers, but open-ended comments were also collected using the comments section on the feedback forms. The responses from the patients were then analysed as specific themes (detailed below).

Table 3: Eligibility criteria for patient focus group

Men and women
Have never previously had a joint replacement operation
Listed for a primary THR/TKR
Good command of English
Any age (18 years and over)
Who agree to partake in focus group and follow up

Four themes were identified for discussion in the patient focus group.

1. Importance of the proposed research study
2. The use of accelerometers in the study (described as “activity monitors”)
3. Questionnaires and the potential burden on patients
4. Potential interventions to improve patient outcomes

Participants of the focus group were also asked to provide feedback about the focus group and how it was conducted.

3.3 Results

3.3.1 Patient demographics

There were two men and three women. Three patients were awaiting a TKR and two patients a THR. The age range was 53 to 84 years of age.

3.3.2 Theme 1: Is this study important for patients?

All 5 participants agreed that this study was important for patients. There was broad agreement that this study was important and would benefit patients undergoing a total hip or knee replacement. In particular, there was consensus that identifying the significant minority of patients who were likely to have a poorer outcome than expected was very important in order to instigate targeted therapies to help improve their outcome.

Participant 1: "*I hadn't realised so many patients were not happy after their knee replacements. If your study helps in working out why this happens, then this would be a good thing"*

Participant 3: *"anything that can help so many patients to do better after their operation is worth doing"*

Participant 4: *"I have friends and neighbours who have had a hip and knee replacement and some of them have not been very happy after their operation. If the study can help provide more information about why this is the case, then this is a very important study to conduct. In fact, I was almost put off wanting to go ahead with my knee replacement because I saw how much pain my neighbour was in after her knee replacement!"*

3.3.3 Theme 2: Explore the acceptability of using accelerometers (activity monitors)

The participants were asked the following questions in relation to the use of accelerometers in the study.

What do you think of them?

All participants felt that the use of accelerometers in the study was an extremely useful aspect of the study and their use in this study was a very interesting idea.

Participant 2: *"This is such an interesting gadget and I like the idea that this device helps record my activity"*

Participant 3: *"This is just like those step counters you can get for free in McDonalds!"*

Would you wear one on your hip during the day?

All participants were happy to wear the accelerometers during the day.

Participant 5: *"Would I need to remove the monitor when I went swimming? Is the monitor waterproof?"*

It was explained to the participants that although the monitors were splash-proof, patients were advised that they should remove the monitors when swimming or having a shower or bath.

Would you be happy to wear one before and after your operation, for one week at a time?

All participants replied were happy to wear the accelerometers at the timepoints specified in the study protocol and were all happy to return the accelerometers in a prepaid envelope.

Are they too heavy?

All participants stated that the accelerometers were not too heavy, and everyone thought that they were very light and comfortable to wear and unobtrusive.

Participant 1: *"it really is very light and comfortable. It's just like wearing a belt"*

Do you feel comfortable wearing them?

All participants stated that they were comfortable to wear.

Participant 4: *"Do I need to wear the monitor next to my skin or can I wear it on top of my clothing?"*

Are there any problems with the accelerometer?

In general, the participants did not have any significant concerns about the accelerometers. Some participants did make the following comments:

Participant 2: *"it's important that the straps are fitted onto the activity monitors before they are given to the patient* "

Participant 5: *"it is important to have a member of the research team help fit the activity monitors onto the patient's waist when they are given the monitor for the first time and to go through the instructions together as well as providing an information leaflet"*

Participant 3: *"The activity monitor won't really pick up activities like cycling as you won't be moving your waist very much? Or swimming as you can't wear them in the water.*

Do you think other patients would also be happy to wear them as part of the study?

All participants stated that they thought patients would be happy to wear them as part of the study.

Do you understand the instructions provided for wearing them?

All participants agreed that the instructions were clear. The participants did state that it would be helpful for a member of the research team to initially demonstrate how to wear the accelerometers.

Are there any other comments you would like to make regarding the accelerometers?

Participant 4: *"the results will be very interesting to analyse the progress of patients after their operation to see how they do"*

Participant 3: *"this is such a novel way of looking at someone's activity level and makes the study really interesting. I would like to see my own activity monitor report if I was taking part in the study"*

3.3.4 Theme 3: Questionnaires and the potential burden on patients

Do you think there are too many questionnaires to complete?

There were four participants who did not feel that there were too many questionnaires to complete but one participant stated that the number of questionnaires could appear quite formidable at first sight.

What is the maximum number of questionnaires you would be happy to complete, if you were part of this study?

The participants felt that between six to twelve questionnaires was the maximum number of questionnaires to complete at each time point of the study. One participant was happy to complete any number of questionnaires as long as there was time permitted. It was noted by one participant that a good time to send out the questionnaires was pre-operatively and at four weeks post-

operatively, which would give patients two weeks to complete them and return them at their six-week follow up clinic appointment.

Participant 1: *"Patients are often limited in what activities they can do during this time period and therefore would have more time to complete the questionnaires."*

Participant 2: *"Patients often feel involved and are likely to want to complete the questionnaires as they feel they are being "looked after" even after their discharge from hospital"*

Have you had any difficulty in completing the questionnaires?

All participants stated that they did not have any difficulty completing the questionnaires. Three of the participants stated that some of the questionnaires had small print font size.

Participant 5: *"Some of the forms have small sized text and could make completing the questionnaires more difficult for elderly patients or those who had a visual impairment"*

Participant 1: *"It's important that the questionnaires are provided in an easy to read format"*

If so, which questionnaire(s) have been difficult to complete?

The BPI questionnaire was felt to have too small a print size. The best laid out questionnaires were identified as being the OHS/OKS and HADS/EuroQoL and SAPS.

How long has it taken you to complete each questionnaire?

All participants took between five and ten minutes per questionnaire to complete. All participants agreed that a two-week time period to answer the questionnaires seemed a good way to ensure patients had time to complete them.

Are there any questions that are difficult to understand?

All of the patients stated that none of the questionnaires were said to be difficult to understand.

Would you prefer to complete the questionnaires at home or during your clinic appointment, whilst waiting in the waiting room?

Four of the five participants preferred to complete the questionnaires at home. One participant preferred to complete the questionnaire during their clinic appointment.

Participant 3: *"You are often waiting around for a long period of time waiting to see the doctor so filling in the forms in clinic seems a good time to do so"*

Would you be happy to complete some of the questionnaires by post?

All participants were happy to complete the questionnaires by post.

3.3.5 Theme 4: Post-operative rehabilitation and care and patient expectations

The participants raised a number of issues with regards to the post-operative rehabilitation support that patients receive after hip and knee replacement surgery. There was considerable variation in the level and intensity of physiotherapy support patients received after their operation. Some patients felt that their physiotherapy support was fine whilst others felt that they had minimal physiotherapy input and that this hindered their recovery and rehabilitation. Most felt that face-to-face physiotherapy was necessary in order to demonstrate the exercises needed to help with their rehabilitation. Pain management post-operatively was a major issue for all patients with the level of pain after the operation often exceeding the patients' expectations of the amount of pain they would be in. Patient expectations were an important element in determining how satisfied patients were likely to be after their

operation. Function as well as pain relief was seen as important outcomes following surgery.

Patient feedback on the conduct of the focus group

The table below (Table 4) shows is a summary of the feedback received from the patient members of the focus group. In summary, the focus group was well received, and the feedback was very positive.

Table 4: Summary of participant feedback on the focus group

Question	**Yes (%)**	**No**
The focus group was better than I expected	6 (100)	0
The topics discussed were interesting	6 (100)	0
The questions were easy to understand	5 (83)	1
I enjoyed discussing this topic with the other participants	4 (67)	2
We were given enough time for discussion	6 (100)	0
The facilitators encouraged participation	6 (100)	0
I got a chance to have my say	6 (100)	0
I felt I was listened to	6 (100)	0
A focus group is a good way of consulting with patients	6 (100)	0

3.4 Discussion

The patient focus group provided patient-centred evidence that the research question being proposed for POSt was relevant and of clinical importance. The outcome of the focus group also helped with the design and layout of the patient questionnaires, which was especially important as POSt was a very questionnaire intensive study and a potential concern was one of questionnaire fatigue amongst study participants, given the number of questionnaires they were being asked to complete. Reassuringly, the focus group participants provided reassurance that this was unlikely to be the case but one of the objectives of the pilot study would be to identify if this was an issue. The focus group provided helpful suggestions on the language and layout of patient facing documents such as the PIS and consent forms.

The use of accelerometers in the study proved to be of great interest to the patient focus group participants, and they found the accelerometers to be comfortable to wear, discrete and easy to take on and off. The patients' views on the use of accelerometers is one of the first examples of obtaining qualitative data on the impact of accelerometers and there is growing evidence that such technology could be used in the future to follow patients up remotely after hip and knee replacements. As a result of the focus group discussion, the study protocol was modified in the following manner, for patients to wear and look after the accelerometers.

- a member of the research team would show patients how to wear and look after their accelerometers as well as providing a patient information sheet on the accelerometers.
- the accelerometers could be returned by pre-paid envelopes back to the research office.

The potential burden on patients completing a large number of questionnaires was explored extensively during the focus group. The issue of questionnaire burden in clinical research studies has been recognised before(Rolstad et al. 2011). Increased response burden has been associated with lower response

rates, completion rates and poorer data quality(Snyder et al. 2007). In the latest 2017/18 Patient Research Experience Survey by the NIHR, one of the reasons patients provided for not wanting to take part in further studies was that there was an issue with the research process and that there were too many questions to answer(Golsorkhi & Steel 2018). One participant in this survey stated that *"sometimes I find there are too many questions and too many of them are on the same subject"*. It was reassuring that the group consensus was that the number of questionnaires being used in the study was not too excessive and that patients would have the time to complete them. The following changes to the study protocol were made.

The OHS/OKS questionnaire was clarified so that forms referred to the operated or affected side (i.e Right or Left Hip or Knee respectively).

- Large print format for the questionnaires was used to help patients to read them clearly.
- Patient anonymity was ensured when sending out the questionnaires by post by the use of a designated patient number.
- If patients had difficulties with some of the questions, a research team member would contact them to clarify and help the patient complete the missing question(s).

Finally, the focus group provided some insight into some of the possible factors that may influence patient dissatisfaction after hip or knee replacements and patient expectation was identified as being a significant factor. Post-operative pain was also a significant issue and often exceeded the patient's expectation of how much pain would be present after surgery. This finding is supported by numerous studies into the role of pain in determining patient satisfaction after THR and TKR(Baker et al. 2007; Lavand'homme & Thienpont 2015; Dunbar et al. 2013; Ali et al. 2014; Brander et al. 2003). There was a wide variation in the amount of physiotherapy support patients received after surgery and all of the focus group participants felt that face to face physiotherapy was required in order to demonstrate to patients the physiotherapy exercises required. The variation in physiotherapy practice has been associated with patient

dissatisfaction after TKR(Johnson et al. 2013). Furthermore, such variation in practice can lead to sub optimal outcomes in terms of muscle strengthening after joint replacement(DiRusso 2014). The lack of physiotherapy support was felt to be a hindrance to rehabilitation and recovery after surgery. All of the participants agreed that improvement in function was equally important as relieving pain and that both were essential outcome measures following hip or knee replacements.

The feedback from the focus group participants provided strong support for the use of focus groups in gaining patients; perspectives on the research study, helping to shape and modify the study protocol, thus ensuring that the study remained relevant and important for patients. Furthermore, the feedback provided valuable information on how best to conduct any further focus groups for future research studies.

3.5 Conclusion

The patient focus group was important in helping to design and modify the POSt protocol, considering the suggestions provided by the focus group, demonstrating all of the benefits of involving PPI at the very start of the research study.

Chapter 4 Pilot study and general feasibility

4.1 Introduction

A pilot study is an integral part of a research study. A pilot study can be defined as “a small-scale test of the methods and procedures to be used on a larger scale…”(Porta et al. 2014). In essence, a pilot study assesses the feasibility of running a larger study and pilot studies have inherent strengths and limitations that must be acknowledged by the research team. A well-conducted pilot study provides information on patient recruitment and retention, adherence to the research protocol, managing research data and enables the researcher to develop a data analysis plan that can be used with the larger study(Arain et al. 2010). Pilot studies can provide information on the day-to-day practicalities of running a research study, identifying areas for improvement and make recommendations for the larger study including creating a data analysis plan using pilot data for future use(Lancaster et al. 2004). Pilot studies cannot however use inferential statistics and create p-values using pilot data. This is because pilot studies are not sufficiently powered to enable appropriate statistical inference to be made nor do pilot studies provide data to estimate sample size for the larger study(Leon et al. 2011).

There are two types of pilot studies. An external pilot study is conducted independently from the definitive study and is primarily conducted when planning for a randomised controlled trial (RCT). An internal pilot study is incorporated as part of the main study design and the data from the pilot may be used (with caution) as part of the overall analysis of the main study(Lancaster 2015). Conducting a pilot study can be advantageous for the following reasons. The data from the pilot study can help assess the feasibility of the main study. It can help explore how well a study protocol is adhered to and accepted by participants, provide information on potential harm to patients, maximise the efficiency of the study and develop methods for handling the data from the main clinical trial(Loscalzo 2009). A pilot study can accomplish all of these objectives with relatively few patients and with limited resources. Feasibility of recruitment and retention are all valuable objectives of a pilot study(Leon et al. 2011).

There are several limitations in terms of what a pilot study can reasonably deliver, and it is critical that the objectives of the pilot study are clear and precise. A pilot study is not hypothesis testing and inferential statistical analysis should be avoided. Thus, tests of significance (p-values) should not be provided by the pilot study(Leon et al. 2011). The use of pilot studies to provide data for sample size calculations is contentious. Some authors state that a pilot study should not be used to determine sample size for the main study because the pilot study population is small and can lead to imprecise sample effect sizes. For example, a pilot study sample size effect which is too great (more false positives) can lead to the main study being under-powered. If the pilot study sample effect size is too small, then there may be more false negatives and the outcome of interest, even if relatively common, may be missed(Leon et al. 2011; Loscalzo 2009). Other studies suggest that sample size calculations is one of the primary purposes of undertaking a pilot study especially when designing a RCT(Lancaster et al. 2004). The flowchart and checklist below provide a framework in terms of identify which areas for amendment are required, which procedural changes are acceptable when transitioning from the pilot study to the main study whilst ensuring the integrity of the methodology is maintained and whether or not to use the pilot study data as part of the main study data analysis(Charlesworth et al. 2013).

4.2 Objectives

The primary objectives of the pilot study were:

1. To assess the general feasibility of the study in terms of consent, recruitment and retention rates.
2. To determine the burden of questionnaires and use of accelerometers for participants.
3. To assess data completeness for the questionnaires and accelerometers.
4. To identify if there were any problems with the design and methods of the study.

The data from the pilot study may help reduce the burden of questionnaires on patients in the larger study, by providing further information on the use of the questionnaires and whether or not certain questionnaires could be removed and still achieve the stated primary objective of the study. The data will also help identify any potential flaws

in the process of conducting the study, which could then be addressed before proceeding to POSt Main. The data from the pilot would also strengthen any future grant application for the POSt-Main study.

The primary outcome measure was the Oxford Hip or Knee score and Patient satisfaction post operatively at 6 months. A poor outcome was defined as an OHS or OKS of equal to or less than 27 or a SAPS outcome of dissatisfaction (Very or Quite dissatisfied).

Figure 1: Flowchart of decision points in a pilot trial(Charlesworth et al. 2013)

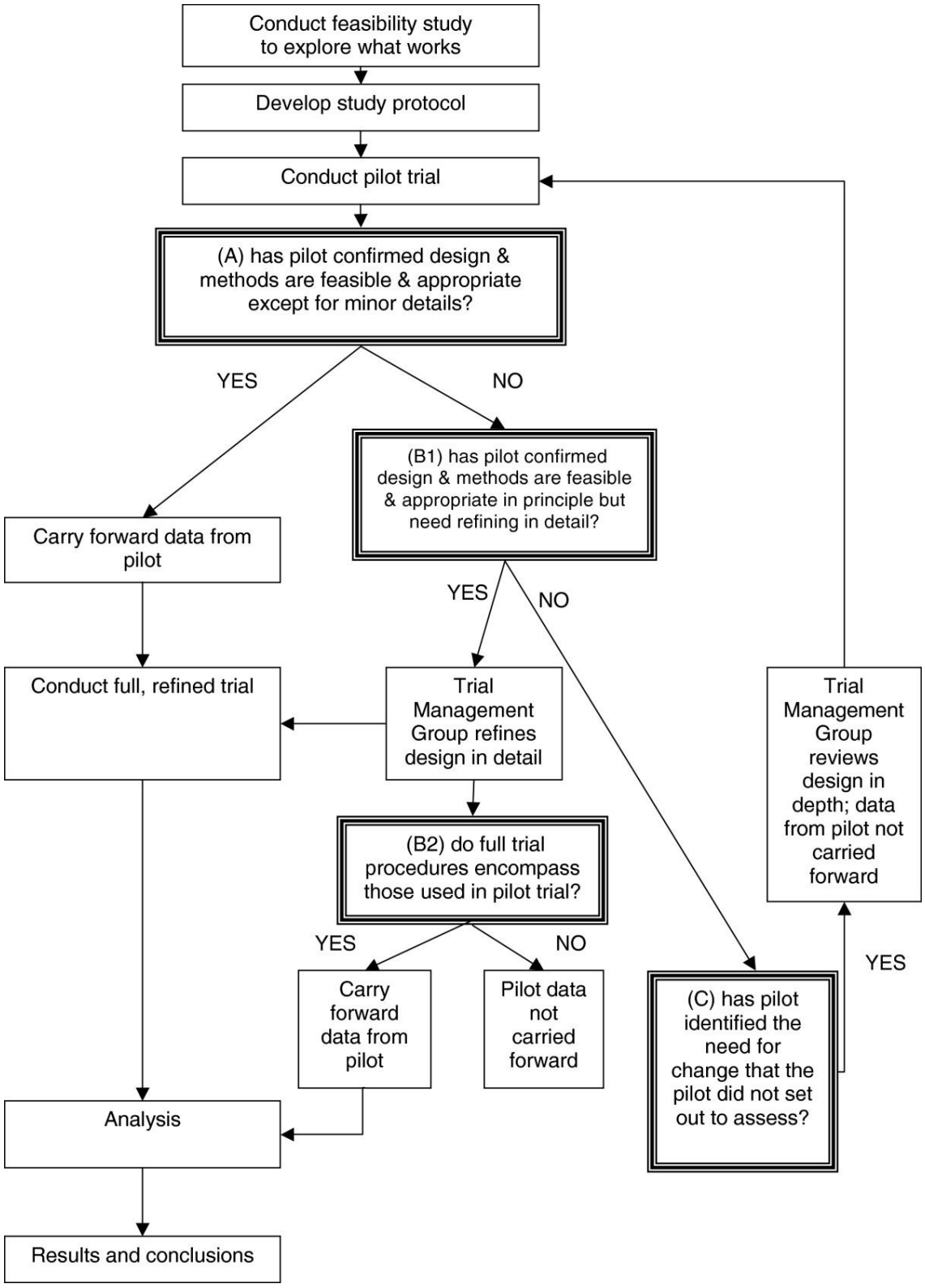


Table 1: Acceptance checklist for clinical effectiveness pilot trials (ACCEPT): trial components, exemplar monitoring methods and exemplar outcomes(Charlesworth et al. 2013)

Component of trial		Monitoring methods (exemplars)	Amend?	Outcomes (exemplars)
Trial design		Review research protocol especially balance of scientific & practical needs	Yes/No	Amend trial design & dependent components. Submit amendment to Research Ethics Committee
Sample size		Test assumptions within protocol on: number of (active) centres; recruitment rates; retention rates; & SD of primary outcomes	Yes/No	Revise if necessary: sample size calculation; trial period; & funding
Interventions	Clinical governance	Assess compliance with: formal training in intervention; Health & Safety regulations; & other clinical governance requirements	Yes/No	Enhance formal training of intervention providers
	Intervention fidelity	Measure & assess adherence to intervention manual by video, observation or audio	Yes/No	Enhance clinical supervision of intervention providers
Participants	Recruitment strategy	Assess: flows of participants; cost & productivity of each route	Yes/No	Refine recruitment strategy, generally & locally
	Eligibility criteria	Assess: characteristics of sample; barriers to recruitment; update of intervention	Yes/No	Refine eligibility criteria
Consent procedures	Participant Information Sheets (PIS)	Consult participants & refusers	Yes/No	Refine PIS especially to address frequently asked questions
	Taking informed consent	Audit consent documentation. Measure & assess adherence to consent procedures by video, observation or audio	Yes/No	Enhance training of research team
Randomisation process		Check quality especially: accessibility by researchers; validity of CONSORT flowchart; & accuracy of stratifying variables	Yes/No	Refine: randomisation procedure & parameters; & training of research team
Blinding		Check whether assessors can predict individual allocations. Test whether unblinded researchers can keep other researchers blind	Yes/No	Refine blinding procedures, e.g. by reallocating responsibilities within research team
Data	Data collection	Assess adherence to interview schedules & fieldwork handbook, including duration of assessments, by video, observation or audio	Yes/No	Refine schedules to reduce assessment burden. Enhance training of research team
	Data quality	Test missing data procedures within draft analysis plan	Yes/No	Refine data collection tools & missing data procedures
	Data management	Test trial database, related procedures & link to analytical software	Yes/No	Refine trial database & procedures
Research Governance	Research protocol adherence	Enable quality assurance officer (QAO) to test adherence as widely as possible	Yes/No	Refine: protocol; quality assurance plan & training of team
	Adverse events (AE)	QAO to test procedures for: reporting AEs; assessing severity, causality & expectedness monitor at management group; report to DMEC	Yes/No	Refine AE reporting & assessment procedures
	Health & Safety	Test H&S procedures, e.g. for lone working	Yes/No	Refine H&S procedures
Data analysis		Test draft analysis plan on pilot data	Yes/No	Refine analysis plan to address research aims in full
Trial management		Review role descriptions of research team. Review remits of trial management group, trial research team etc.	Yes/No	Review role descriptions of research team. Refine roles e.g. if workloads vary. Refine remits e.g. if inadequate reporting of any component

4.3 Methods

As preparation for the pilot study, the necessary patient facing documents such as the patient information sheets (PIS) (Appendix 13) and consent forms (Appendix 14) as well as the patient study packs and case report files (CRF) were prepared with advice provided by the Chair of the Patient Liaison group and based on the outcome of the patient focus group. The local Research Ethics Committee (REC) application was then submitted to the East Midlands Research Ethics Committee and was given a favourable opinion. Posters, highlighting the research study, were placed in outpatient clinic areas (Appendix 1).

4.3.1 Screening and eligibility

Using the waiting list register, patients who were listed for a primary hip or knee replacement were identified as potential participants. Recruitment of potential participants involved the screening of medical notes in order to identify if potential participants were eligible to take part in the study based on the inclusion and exclusion criteria as detailed before (see Figure 2).

Patients who met the inclusion criteria were then sent patient information sheets (Appendix 2) detailing the exact details of the study. I or another designated member of the research team approached potential participants in the pre-admission clinic in order to recruit potential participants. Valid informed consent (Appendix 3) from patients who agreed to participate was then obtained. For the pilot study, the first 100 patients were prospectively recruited. Five patients had their operations cancelled or postponed thus leaving 95 patients recruited to the pilot study. A summary flowchart of the POSt Pilot study is shown in Figure 3.

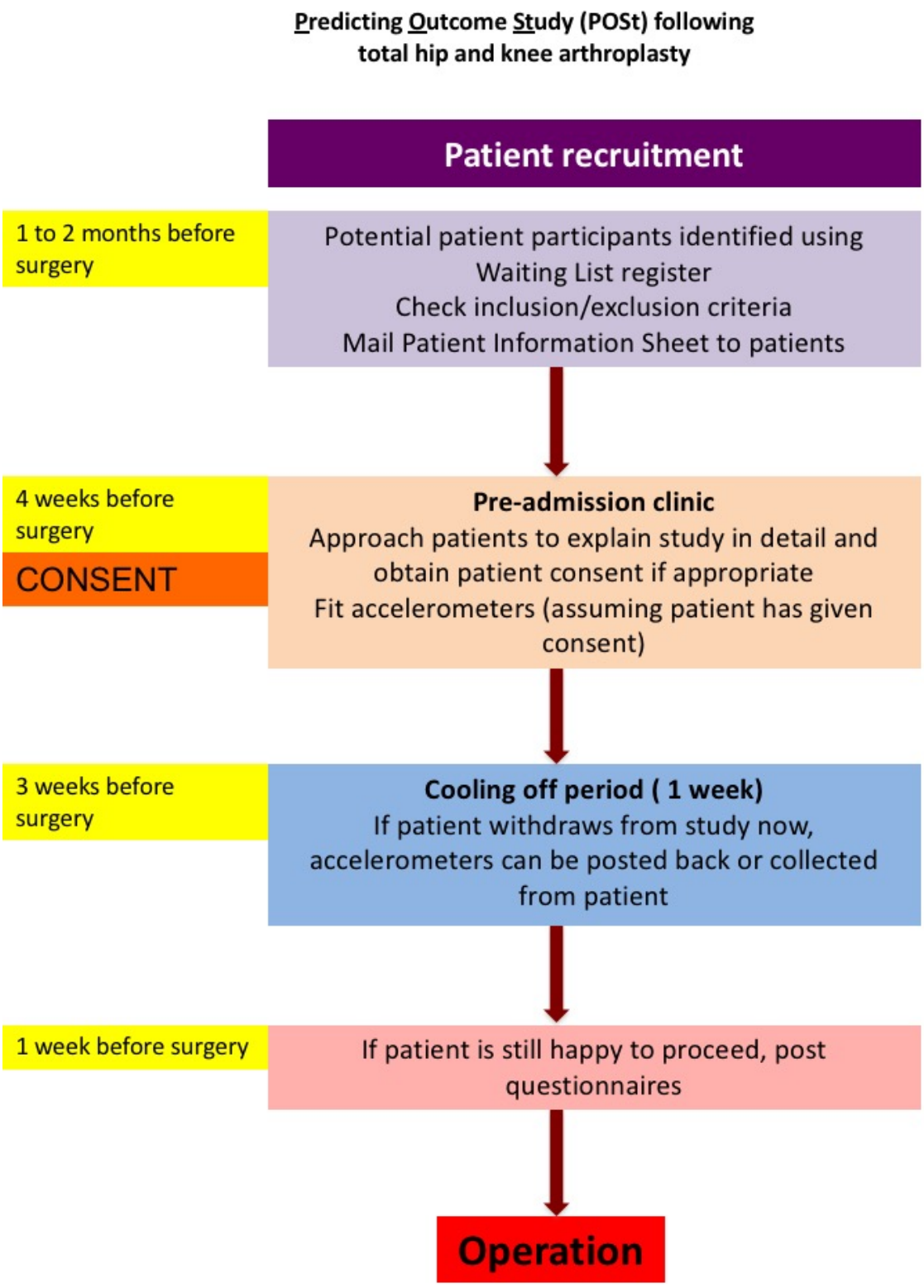
Inclusion criteria:

- Osteoarthritis of the hip or knee;
- Able to give informed consent;
- Able to read and write English;
- First hip or knee primary arthroplasty

Exclusion criteria:

- Other types of arthritis/Inflammatory arthritis (Rheumatoid Arthritis; Psoriatic arthritis; Osteonecrosis; Post-traumatic arthritis)
- Revision arthroplasty;
- Previous hip or knee arthroplasty;
- Vulnerable adult groups (patients with learning difficulties, dementia, drug and alcohol problems, significant mental health problems)

Figure 2: Flowchart of patient recruitment



4.3.2 Feasibility

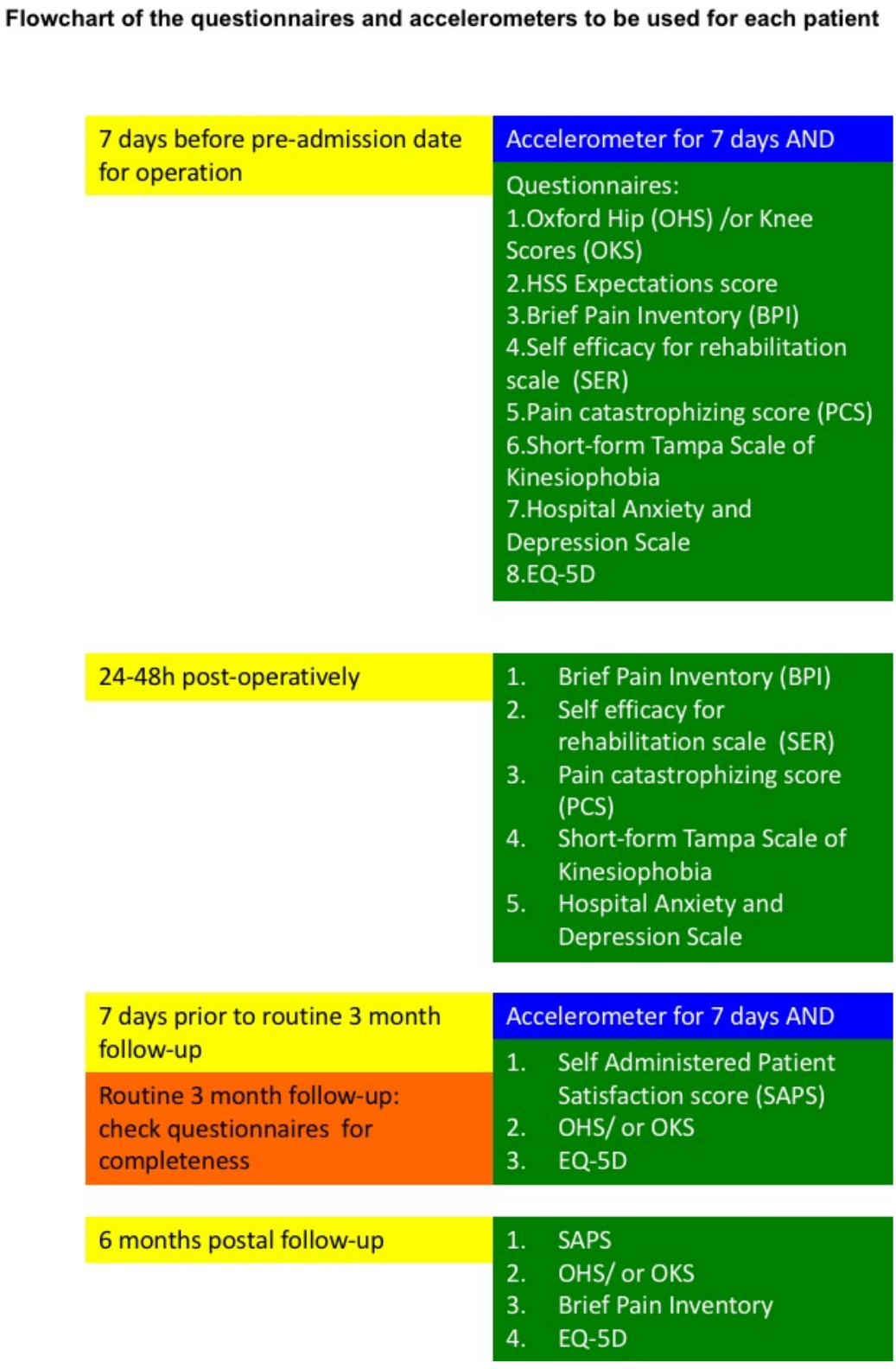
A sample size calculation was not performed for the pilot study on its own but rather for the main study (POSt Main) as this was intended to be an internal pilot and the data from this pilot would also be included in the final data analysis of the main study. However, the number of participants (n=100) for the pilot seemed a reasonable number and proportion (25%) of the total number of participants for the main study based on the literature, in order to provide useful information on recruitment, feasibility and the other objectives of the pilot study (Machin et al. 2018, Hoboken et al. 2016). This was also the sample size recommended by the APG (Advanced Postgraduate) review committee confirming my progression from APG to full PhD status, suggesting that the PhD thesis be based on the pilot study using a sample size of approximately 100 patients.

For the main study (POSt Main), a sample size calculation was performed to allow for a dropout rate of up to 25%. The minimum sample size was 150 patients in each of the two medical conditions (hip or knee replacement group) which would give the ability to detect an odds ratio of 3 for a poor outcome, with an 80% power and significance of 0.05 for a risk factor (such as pain catastrophizing) with a prevalence of 5% in the study sample. In our centre, 650 THAs and 750 TKRs are performed annually so the POSt Main study was deemed feasible in terms of data collection and patient recruitment within a year. The POSt Main study aimed to recruit 400 patients with 200 patients in the hip and knee group respectively. The predicted recruitment rate was 4 patients/week. The feasibility of POSt-Main was determined by analysing data from POSt Pilot on consent, recruitment and retention rate and by analysing the amount of missing data present in the questionnaire and accelerometer data at each timepoint.

4.3.3 Questionnaires

All patients were given a number of questionnaires and were asked to wear the accelerometers at different timepoints throughout the study. The study flowchart is shown in Figure 2.

Figure 3: Flowchart of the questionnaires and accelerometers to be used for each patient.



The questionnaires (see Appendices 5 to 15) sent to patients are detailed below and were designed to assess a range of factors and outcome measures, including physical, psychological and functional variables.

1. Pain Catastrophizing Score (PCS)
2. Short-form Tampa Scale of Kinesiophobia-11(TSK-11)
3. The Brief Pain Inventory (BPI)
4. Hospital Anxiety and Depression Scale (HADS)
5. Self-efficacy for rehabilitation scale (SER)
6. Oxford Hip (OHS) and Knee Scores (OKS)
7. EQ-5D
8. Self-administered Patient Satisfaction (SAPS) questionnaire
9. Patient Hip and Knee expectations questionnaires (Hospital for Special Surgery)

Pain Catastrophizing Scale (PCS)

The Pain Catastrophising Scale was developed in 1995 by Sullivan et al. (Sullivan et al. 1995) with the aim of attempting to create an all-encompassing instrument for measuring the various aspects of the phenomenon of pain catastrophizing. It is one of the most widely used measure of pain catastrophizing. It covers the three main domains of rumination, magnification and helplessness.

Short-form Tampa Scale of Kinesiophobia (TSK-11)

The Tampa Scale of Kinesiophobia-11 is a shortened form of the original questionnaire, which is a widely used form designed to assess the fear of pain in patients with back pain(Woby et al. 2005). The TSK-11 version has similar psychometric properties to the original TSK version but with the added advantage of brevity(Woby et al. 2005). The TSK-11 form has been proven to be a brief, valid and reliable measure of pain-related fear of movement in chronic pain patients(Tkachuk & Harris 2012).

The Brief Pain Inventory (BPI)

The BPI is a widely used self-administered questionnaire originally designed to assess pain in cancer patients(Cleeland 1991), but its use has extended to include patients with a variety of different pain inducing pathology including osteoarthritis and musculoskeletal conditions. It is available as the short (9-item) or long (17-item) form(Poquet & Lin 2016). For this study, the short form was used and is the more frequently used version in clinical research.

Hospital Anxiety and Depression Scale (HADS)

The HADS questionnaire is a self-screening, 14-item form consisting of anxiety and depression subsections, used extensively in clinical practice(Stern 2014). It was first devised over 30 years ago by Zigmond and Snaith(Zigmond & Snaith 1983).

Self-efficacy for rehabilitation scale (SER)

The SER is a patient-reported questionnaire designed to assess patients' beliefs about their ability to undertake activities which are routinely used in physiotherapy exercises(Waldrop et al. 2001). The scale consists of 12 items that are added together and converted into a 100-point scale, with a higher score indicating more self-efficacy.

Oxford Hip (OHS) and Knee Scores (OKS)

The OHS and OKS questionnaires are validated, patient self-reported with 12 questions, each with five options scoring from 0 to 4, all related to pain and function(Dawson et al. 1996; Dawson et al. 1998). The worst possible score is 0 and the best possible score is 48(Murray et al. 2007). In the original publications each section scored from 1 to 5 with 1 being the best and 5 the worst (12-60 with 12 being the best and 60 being the worst) but in 2007, Murray et al. recommended changing the scoring system to the current 0-48 range(Murray et al. 2007). The OHS and OKS are the Patient Reported Outcome Measures (PROMs) used in the UK for all primary THR and TKRs as registered in the National Joint Registry (NJR). The decision to use the OHS and OKS in the study was because it is simple to use, well validated, patient (rather than clinician) reported and is the preferred PROM used by the NJR and many

other clinical studies measuring outcome after THR and TKR with an excellent clinical track record.

EQ-5D

The EQ5D is a quality of life assessment, preference-based measure used extensively in assessing health technology. The EQ-5D comprises five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (EQ5D-3L User Group 2013). The EQ-5D is collected together with the OHS/OKS as part of the UK national PROMs survey.

Self-administered Patient Satisfaction (SAPS) questionnaire

The Self-Administered Patient Satisfaction scale was designed in 2011 to assess patient satisfaction after hip and knee replacements(Mahomed et al. 2011). It is scored on a 4-point Likert scale. The response categories are: very satisfied (100 points), somewhat satisfied (75 points), somewhat dissatisfied (50 points), and very dissatisfied (25 points). The scale score is the unweighted mean of the scores from the individual items, ranging from 25 to 100 per item (with 100 being most satisfied).

Patient Hip and Knee expectations questionnaires (Hospital for Special Surgery)

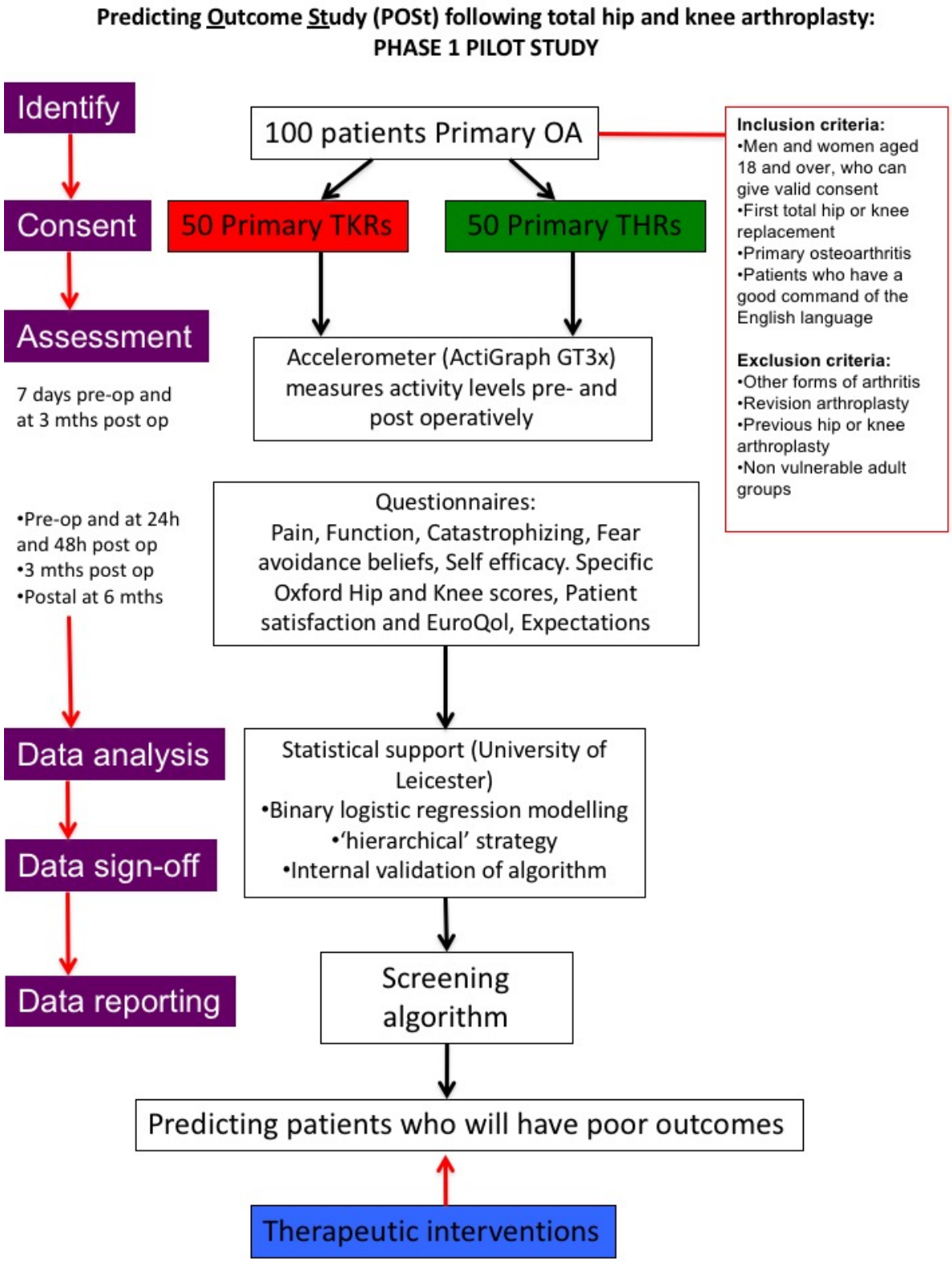
The HSS Hip and Knee Replacement Expectations Survey are 18-item and 19-item surveys respectively and include 5 different categories of expectations: pain, walking, psychological state, essential activities and non-essential activities(Mancuso et al. 1997; Mancuso et al. 2001).

4.3.4 Accelerometry

Accelerometers are devices, which measure the physical activity performed by patients. Patients were asked to wear an accelerometer for 7 days prior to admission and 7 days prior to their routine 3-month postoperative review. Patients were shown how to wear the accelerometer around their waist and were provided with a patient information sheet about wearing the accelerometer (Appendix 4). Physical activity and

standing, sitting, and lying positions were assessed using the tri-axis ActiGraph GT3X Activity Monitor (ActiGraph LLC, Pensacola, FL, USA). The device has dimensions of 1.50 x 1.44 x 0.70 inches and has a weight of 28 g. The Actigraph GT3X was released in 2009 and contains an ADXL335 accelerometer (Analog Devices, Norwood MA). The ADXL335 is a 4 × 4 × 1.45 mm triaxial capacitive MEMS sensor with a full-scale range of ± 3g(JOHN & FREEDSON 2012). ActiLife software (version 5.6.1; ActiGraph LLC, Fort Walton Beach, FL, USA) was used to analyse the data. Data was gathered for 7 days. The minimum wear time was for 10 hours/day for 4 days in order to provide wear time validation. The Actigraph GT3X not only provides information on physical activity but also uses vector magnitude data from three axes (Axis 1 vertical, Axis 2 horizontal, Axis 3 lateral) to provide inclinometer data, assigning a number from 0 to 3 where the position of the individual is defined as: individual not wearing the monitor (number 0), is standing (number 1), lying (number 2), and sitting (number 3). The ActiGraph algorithm classifies counts above 100 to 1951 counts per minute (cpm) as standing. If the counts are below 100 cpm, the algorithm distinguishes between standing or lying or the monitor not being worn(Barwais et al. 2013).

Figure 5: Flowchart for POSt Pilot summarising main features of the study



Each participant had the right to withdraw study at any time. In addition, the investigator could discontinue a participant from the study at any time if the investigator considered it necessary for any reason including:

- Ineligibility (either arising during the study or retrospective having been overlooked at screening)
- Significant protocol deviation
- Significant non-compliance with treatment regimen or study requirements
- Consent withdrawn
- Lost to follow up

Withdrawal of participants from the study would result in any information received up to the date of withdrawal being potentially analysed on an intention to treat basis. The reason for withdrawal was recorded in the CRF.

4.4 Results

4.4.1 Feasibility

In our centre, 650 THAs and 750 TKRs are performed annually so this study is feasible in terms of data collection and patient recruitment within a year. A sample size calculation was performed to allow for a dropout rate of up to 25%. The minimum sample size was 150 patients in each of the two medical conditions (hip or knee replacement group) which will give the ability to detect an odds ratio of 3 for a poor outcome, with an 80% power and significance of 0.05 for a risk factor (such as pain catastrophizing) with a prevalence of 5% in the study sample. This should be achieved comfortably within our total study sample of 400 patients. The predicted recruitment rate was 4 patients/week. The feasibility of the study was determined by analysing data from POSt Pilot on consent, recruitment and retention rate and by analysing the amount of missing data present in the questionnaire and accelerometer data at each timepoint.

4.4.1.1 Consent rate

This was measured as the number of participants who were approached and provided their consent measured against those participants who were contacted and declined to take part in the study. The consent rate was 72.6%.

Table 1: Consent rate of participants

Consented	98 (72.6%)
Declined	37 (27.4%)
Total	135

4.4.1.2 Recruitment rate

Figure 6: Line graph showing recruitment and decline rates for POSt-pilot

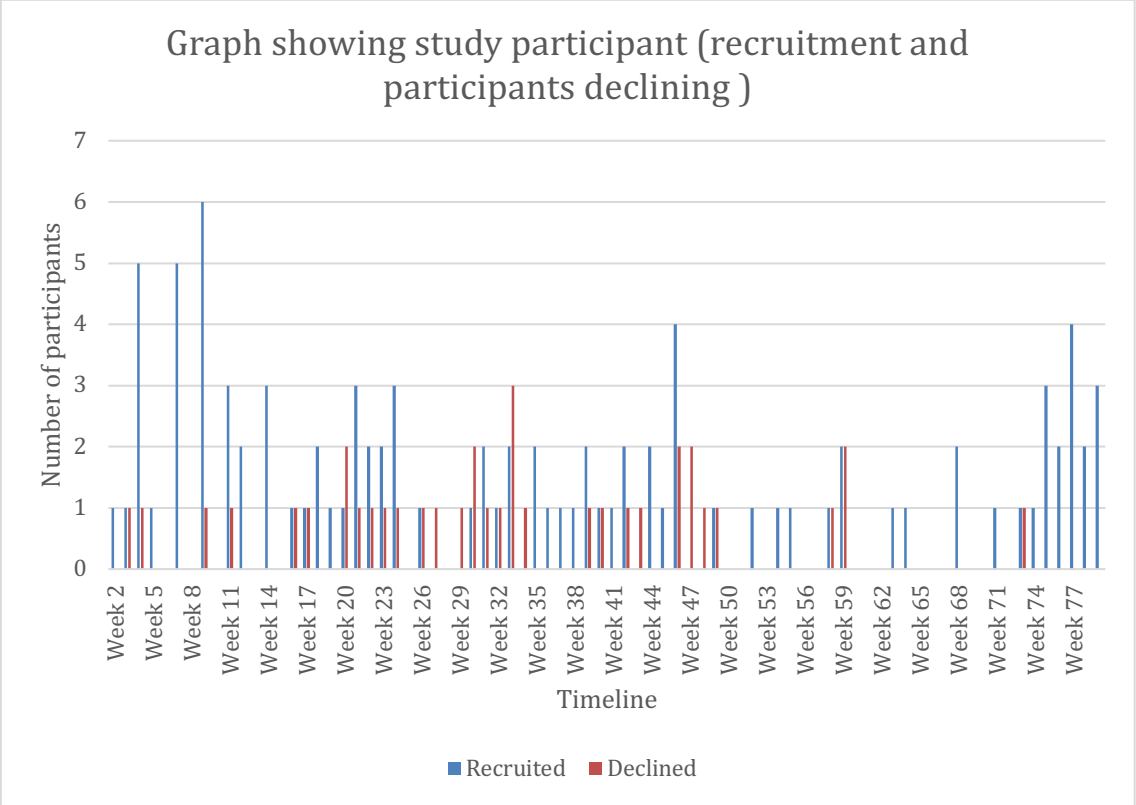
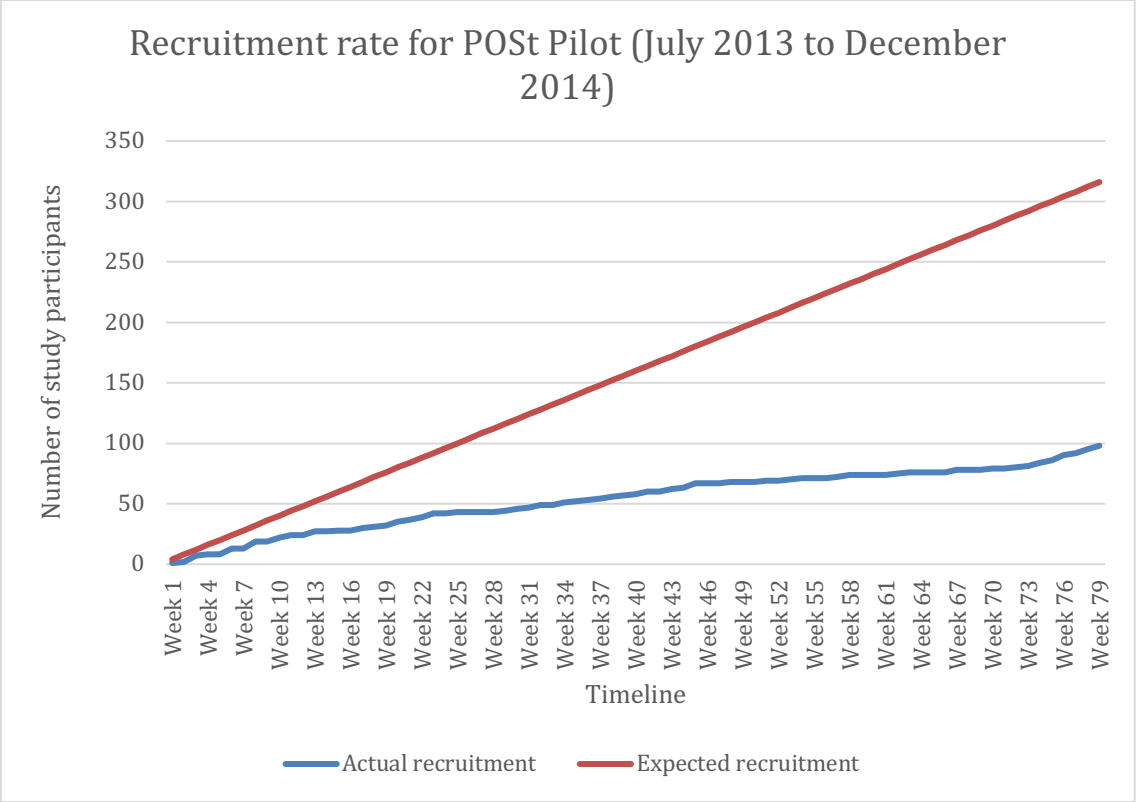


Figure 7: Graph showing expected and actual recruitment rates for POSt-Pilot



The data on recruitment and decline rates show that the recruitment target of 4 participants per week was not reached and by week 79, the recruitment rate was approximately 30% of the expected recruitment rate.

4.4.1.3 Retention rate

Of the 95 patients recruited to the pilot study, 1 patient withdrew and did not complete the 3 or 6-month follow up questionnaires. The retention rate was 98.9%. The reason provided was that the patient felt unable to cope answering the questionnaires due to ongoing knee pain after the knee replacement operation. There were no other withdrawals from the pilot study.

4.4.2 Burden of questionnaires and accelerometers

The questionnaires were well tolerated and only 1 patient commented on the burden of questionnaires being used in the pilot study and withdrew from the study as a result. One morbidly obese patient (BMI >40) was not able to wear the accelerometer as the waist belt was too tight and uncomfortable. Otherwise the accelerometers were well

tolerated. There were 3 other participants who did not wear the accelerometers, but no explanation was provided for this. There were 4 accelerometers lost and were never returned back to the research team. This represented a loss rate of 10% (out of 40 initially purchased).

4.4.2.1 Questionnaire missing data

The amount of missing data in the questionnaires was calculated and the results presented as a series of tables for both hip and knee participants (Tables 2 to 9).

Table 2: Data completeness: Pre-Op timepoint (PR) for hips

Questionnaire	**Complete (%)**	**Incomplete (%)**	**Missing (%)**
Patient demographics			
Age	36 (92.3)	0	3 (7.7)
Gender	39 (100)		
ASA	39 (100)		
BMI	39 (100)		
Comorbidities	39 (100)	0	0
OHS	36 (92.3)	3 (7.7)	0
PCS	33 (84.6)	6 (15.4)	0
SER	33 (84.6)	6 (15.4)	0
HADS	37 (94.9)	1 (2.6)	1 (2.6)
TSK-11	23 (60.0)	10 (25.6)	6 (15.4)
EQ5D	36 (92.3)	0	3 (7.7)
HSS Expectations	31 (79.4)	8 (20.5)	0
BPI	38 (97.4)	1 (2.6)	0

Table 3: Data completeness: Post-Op timepoint (PO) for hips

Questionnaire	Complete (%)	Incomplete (%)	Missing (%)
PCS	35 (89.7)	4 (10.3)	0
SER	31 (79.5)	5 (12.8)	3 (7.7)
HADS	36 (92.3)	2 (5.1)	1 (2.6)
TSK-11	25 (64.1)	10 (25.6)	4 (10.2)
BPI	35 (89.7)	3 (7.7)	1 (2.6)

Table 4: Data completeness: 3-month timepoint (3M) for hips

Questionnaire	Complete (%)	Incomplete (%)	Missing (%)
OHS	31 (79.5)	6 (15.4)	2 (5.1)
SAPS	35 (89.7)	1 (2.6)	3 (7.7)
EQ5D	36 (92.3)	0	3 (7.7)

Table 5: Data completeness: 6-month timepoint (6M) for hips

Questionnaire	Complete (%)	Incomplete (%)	Missing (%)
OHS	28 (71.8)	4 (10.2)	7 (17.9)
SAPS	28 (71.8)	4 (10.2)	7 (17.9)
EQ5D	32 (82.1)	0	7 (17.9)
BPI	31 (79.5)	3 (7.7)	5 (12.8)

Table 6: Data completeness: Pre-Op timepoint (PR) for knees

Questionnaire	Complete (%)	Incomplete (%)	Missing (%)
Patient demographics			
Age	51 (100)	0	0
Gender	51 (100)		
ASA	51 (100)		
BMI	51 (100)		
Comorbidities	51 (100)	0	0
OKS	51 (100)	0	0
PCS	48 (94.1)	1	2
SER	45 (88.2)	3	3
HADS	49 (96.1)	0	2
TSK-11	20 (39.2)	14	17
EQ5D	49 (96.1)	0	2
HSS Expectations	48 (94.1)	9	3
BPI	49 (96.1)	3	2

Table 7: Data completeness: Post-Op timepoint (PO) for knee

Questionnaire	Complete (%)	Incomplete (%)	Missing (%)
PCS	45 (88.2)	4 (7.8)	2 (3.9)
SER	40 (78.4)	9 (17.6)	2 (3.9)
HADS	47 (92.2)	2 (3.9)	2 (3.9)
TSK-11	25 (49.0)	13 (25.5)	13 (25.5)
BPI	47 (92.2)	1 (2.0)	3 (5.9)

Table 8: Data completeness: 3-month timepoint (3M) for knees

Questionnaire	Complete (%)	Incomplete (%)	Missing (%)
OKS	42 (82.4)	4 (7.8)	5 (9.8)
SAPS	45 (88.2)	1 (2.0)	5 (9.8)
EQ5D	46 (90.2)	0	5 (9.8)

Table 9: Data completeness: 6-month timepoint (6M) for knees

Questionnaire	Complete (%)	Incomplete (%)	Missing (%)
OKS	42 (82.4)	3 (5.9)	6 (11.8)
SAPS	43 (84.3)	1 (2.0)	7 (13.7)
EQ5D	45 (88.2)	0	6 (11.8)
BPI	42 (82.4)	2 (3.9)	7 (13.7)

4.4.2.2 Accelerometer missing data

Accelerometer missing data was also calculated and presented in Table 10.

Table 10: Data completeness: Accelerometer data

Accelerometer data	Complete (%)	Incomplete (%)	Missing (%)
Hip PR	36 (92.3)	0	3(7.7)
Hip 3M	33 (84.6)	0	6 (14.4)
Knee PR	45 (88.2)	0	6 (11.8)
Knee 3M	45 (88.2)	0	6 (11.8)

4.4.3 TSK-11 questionnaire

During the pilot study, an error with the TSK-11 questionnaire was identified. An incorrect TSK-11 questionnaire was sent out to 74 participants in the pilot study which mistakenly contained 5 answer columns (“Strongly agree; Agree; Not sure; Disagree; Strongly disagree”. The correct version of the TSK-11 questionnaire only had 4 columns: “Strongly agree; Agree; Disagree; Strongly disagree”. Thus, some participants were able to choose “Not sure” as an option. A potential method for

managing this error is discussed further in Chapter 6 of this thesis. The pilot study helped identify a problem with the TSK-11 questionnaire which could potentially have had a significant negative impact on the main study if this error had not been identified in the pilot study.

4.5 Discussion

This pilot study has demonstrated that the main study, POSt-Main, would be feasible to conduct in terms of consent rates and retention rates. In particular, the consent rate was very good at 72.6% which shows that the majority of patients who were eligible to be included in the study, were happy to take part in the study. Furthermore, the retention rate for participants was excellent at 98.9%.

The recruitment rate was approximately 30% of the expected rate and this is a concern. There are several reasons for this. Firstly, the recruitment was being conducted by only one person (JP) and therefore, many patients who were potential participants of the study were missed because of the difficulty logistically in being in different outpatient pre-assessment clinics at the same time. Secondly, because there was only one research member recruiting, during annual leave, no recruitment took place. Finally, patients were often missed as a result of postponed operations and last-minute cancellations of surgery for a number of reasons including clinical as well as logistical challenges in theatre scheduling. The recommendations going forward to improve the recruitment rate, based on the findings of the pilot study was to employ a dedicated POSt clinical research assistant (CRA) with the support of other CRAs and research administrators from the Clinical Research Network (CRN) so that at any one time, there was at least one CRA available to consent and recruit patients to the study even when research team members were away. This would also allow more clinics to be covered in terms of recruiting potential participants, thereby improving the recruitment rate, whilst also ensuring follow up of research participants was not compromised. The other recommendation to improve the recruitment rate was to ensure that potential participants were contacted by telephone before their scheduled pre-assessment clinic appointment, after being screened for eligibility, in order to

“prepare” the potential participant and provide further verbal information about POSt and to clarify any uncertainties that the potential participants may have about the study. This was in addition to the written information sent out to all potential participants about POSt.

The level of data completeness varied but overall, the data was excellent, especially in the preoperative time period. At the 3 and 6 months follow up, the completeness of the data decreased but still remained high for most of the questionnaires and accelerometer data (for knees, this was >80% and for hips, this was >70%). The exception to this was the data completeness for the TSK-11 questionnaire which was much lower than that for all the other questionnaires. This was the case for both hip and knee patients although the hip patients had a higher level of completeness compared to the knee patients. The reasons for this remain unclear but may reflect difficulty in completing the TSK-11 forms by the study participants. Furthermore, the original TSK-11 form used for the first 74 participants in the pilot study was incorrect and this may have contributed to the increased data incompleteness.

In terms of the burden of the questionnaires, only one participant withdrew from the pilot study for this reason. The accelerometers appeared to be well tolerated with only 4 participants not wearing the monitors and the accelerometer loss rate was acceptable at 10%. The results of the pilot study provide evidence that the burden of questionnaires and use of the accelerometers was well tolerated by study participants and supports the findings of the patient focus group. The loss of accelerometers at 10% was also within the acceptable limit and going forward, provides further information in terms of research costs and incorporating this into future research grant applications.

The pilot study identified a significant problem with the TSK-11 questionnaire with 74 participants being provided with the incorrect version of the form, with an additional Unsure category. A method of mitigating this error was initiated involving the use of the correct and incorrect TSK-11 questionnaire being sent to 21 patients at the pre- and post-operative time points in order to identify any patterns in participants

answering using the Unsure option. This will be discussed further in Chapter 5 of the thesis.

4.6 Conclusion

The pilot study has demonstrated that the main study should be feasible and that the potential burden of questionnaires is minimal and use of accelerometers very acceptable to most study participants. The questionnaire and accelerometer data completeness was excellent and the only main issues raised concerned the use of the TSK-11 and the low recruitment rate but recommendations to manage these challenges should improve the situation.

Chapter 5 Error with the Tampa Scale for Kinesiophobia-11

5.1 Introduction

Kinesiophobia is a pain related fear of movement(Kori et al. 1990). The Tampa Scale for Kinesiophobia (TSK) is a common tool for assessing a patient's pain related fear of movement and was first developed by Miller, Todd and Kori in 1991(Miller et al. 1991). It was initially designed for use in patients with back pain but has been used in other areas of such as in knee replacement patients(Eymir et al. 2019; Güney-Deniz et al. 2017; Filardo et al. 2015; Cai et al. 2018; Sullivan et al. 2009). The original TSK was a 17-item questionnaire with patients rating each item on a 4-point Likert scale with scoring alternatives ranging from 'strongly disagree' to 'strongly agree'. There were four inversely scored items (4, 8, 12, and 16) with the total scores ranging from 17 to 68. The higher the score, the greater the fear of movement or re-injury(Miller et al. 1991; Woby et al. 2005). In 2005, Woby et al. introduced a shortened English language version of the TSK which removed 6 items including all of the inversely scored items and showed similar psychometric properties to the original TSK but with the added advantage of being shorter to complete, more user friendly (from a patient's perspective) and easier to administer(Woby et al. 2005; Larsson et al. 2014). This 11-item TSK version was subsequently called the TSK-11 and has been used in other studies, confirming that the TSK-11 is a brief, validated and reliable measure of pain related fear of movement in chronic pain patients(Tkachuk & Harris 2012; Larsson et al. 2014).

The incidence of kinesiophobia has been shown to be range from 25% to 93% post-operatively in patients who have had a total knee replacement (TKR)(Cai et al. 2018), (Kurtulus et al. 2017). There has been almost no research into the incidence of kinesiophobia in hip replacement patients. Kinesiophobia is one of the main psychological traits that is now thought to be a key predictor of how patients recover

after TKR(Filardo et al. 2015) and is associated with the development of chronic pain and dysfunction after TKR(Heuts et al. 2004). Patients exhibiting kinesiophobic traits who have undergone a TKR have been shown to have a poorer early functional outcomes(Güney-Deniz et al. 2017) with inhibited functional recovery as a result of activity avoidance during the rehabilitation phase after surgery(Doury-Panchout et al. 2015). The use of the TSK-11 was an important aspect of the pilot study in order to collect data on a potentially important factor that may help predict poor outcome after TKR or indeed THR.

When using questionnaires, it is critically important that the correct version of the questionnaire is used(Boynton & Greenhalgh 2004). The content, structure but also the presentation of the questionnaire have been shown to be critically important to how participants answer the questionnaire(Nieuwenhuijsen 2005) and any changes to the layout, design or options for answering the questions will potentially invalidate an already establish validated questionnaire. The POSt-Pilot study identified an error in the use of an incorrect version of the TSK-11 questionnaire. The research aim of this chapter was to explore what steps could be taken to mitigate against this type of questionnaire error including exploring if an appropriate statistical conversion method could be used in order to salvage the collected data for further analysis. The objectives of this chapter was to describe the pattern of responses on the correct questionnaire in people who used the unsure category on the incorrect version of the TSK-11 form and make recommendations in how to avoid making such an error in the first place and how to mitigate against such an error if this occurs.

The reason for undertaking this work was because this was an internal pilot and it was important to identify why the error took place in the first place but also to determine if there were any methods of mitigating against the potential loss of data that would have happened if the error had not been identified as the data from the pilot was being included in the data analysis of POSt Main.

5.2 Methods

During the pilot study, it was identified that the TSK-11 questionnaire that was being used for the first 53 patients recruited to the pilot, was incorrect (See Figure 1). This version of the TSK-11 questionnaire had an additional middle response option of "Unsure" as part of the Likert scale of responses. Instead of having 4 responses (Strongly Agree, Somewhat Agree, Somewhat Disagree, Strongly Disagree), this incorrect version had 5 possible responses (Strongly Agree, Somewhat Agree, Unsure, Somewhat Disagree, Strongly Disagree). The correct version of the TSK-11 questionnaire is shown in Figure 2.

Figure 1: Incorrect Version of TSK-11 with the additional Unsure column (highlighted in red)

Tampa Scale of Kinesiophobia-11 (TSK-11) (Incorrect)

Read the statements and put a tick in the box which reflects your agreement with the statement:

	Strongly Agree	Somewhat Agree	Unsure	Somewhat Disagree	Strongly Disagree
I'm afraid that I might injure myself if I exercise	☐	☐	☐	☐	☐
If I were to overcome it, my pain would increase	☐	☐	☐	☐	☐
My body is telling me I have something dangerously wrong	☐	☐	☐	☐	☐
People aren't taking my medical condition seriously enough	☐	☐	☐	☐	☐
My accident has put my body at risk for the rest of my life	☐	☐	☐	☐	☐
Pain always means I have injured my body	☐	☐	☐	☐	☐
Simply being careful that I do not make any unnecessary movements is the safest thing I can do to prevent my pain from worsening	☐	☐	☐	☐	☐
I wouldn't have this much pain if there wasn't something potentially dangerous going on in my body	☐	☐	☐	☐	☐
Pain lets me know when to stop exercising so that I don't injure myself	☐	☐	☐	☐	☐
I can't do all the things normal people do because it's too easy for me to get injured	☐	☐	☐	☐	☐
No one should have to exercise when s/he is in pain	☐	☐	☐	☐	☐

Figure 2: Correct version of TSK-11 without the Unsure column

Tampa Scale of Kinesiophobia-11 (TSK-11) Correct

Read the statements and put a tick in the box which reflects your agreement with the statement:

	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
I'm afraid that I might injure myself if I exercise	☐	☐	☐	☐
If I were to overcome it, my pain would increase	☐	☐	☐	☐
My body is telling me I have something dangerously wrong	☐	☐	☐	☐
People aren't taking my medical condition seriously enough	☐	☐	☐	☐
My accident has put my body at risk for the rest of my life	☐	☐	☐	☐
Pain always means I have injured my body	☐	☐	☐	☐
Simply being careful that I do not make any unnecessary movements is the safest thing I can do to prevent my pain from worsening	☐	☐	☐	☐
I wouldn't have this much pain if there wasn't something potentially dangerous going on in my body	☐	☐	☐	☐
Pain lets me know when to stop exercising so that I don't injure myself	☐	☐	☐	☐
I can't do all the things normal people do because it's too easy for me to get injured	☐	☐	☐	☐
No one should have to exercise when s/he is in pain	☐	☐	☐	☐

The incorrect and correct TSK-11 questionnaires were sent to the next 21 patients pre-operatively and 19 patients post-operatively (between 24 and 48h after surgery). The incorrect TSK-11 questionnaire was placed at the start of the pack of questionnaires that patients needed to complete as part of the POSt Pilot study. The correct version was placed at the end of the pack. In this way, it was hoped that the interval between completing the incorrect and correct TSK-11 would help mitigate against the tendency for patients to be influenced by how they answered the incorrect TSK-11 questionnaire.

In order to accomplish this, patients were asked to complete both a correct and incorrect version of the TSK-11 in order to examine the data for patterns in how patients responded. If a pattern could be observed, then it might be possible to determine if there was a way of statistically determining if an Unsure (U) response would lead to a particular different response. For example, if all Unsure responses

were subsequently Strongly Agree responses, then it would be reasonable to conclude that all of the Unsure responses could be regarded as being Strongly Agree.

5.3 Results

There were 74 patients analysed in this dataset. There were 37 hip replacement patients and 37 knee replacement patients. There were 20 patients who were sent both the correct and incorrect versions of the TSK-11 questionnaire. In this cohort, there were 8 patients in the hip group who were sent pre-operatively both versions of the TSK-11 questionnaires and there were 12 patients in the knee group. Post-operatively (between 24 and 48h after surgery), eight of the same original nine hip patients who were sent both versions of the TSK-11 questionnaires, completed both versions. In the knee group, 11 patients completed both versions of the TSK-11 questionnaire post-operatively. In the knee group, five patients completed both versions pre-operatively but not post-operatively (JP065, JP071, JP072, JP077, JP087). Four patients in the knee group completed both versions post-operatively but not pre-operatively (JP076, JP086, JP088, JP094). Seven patients completed both versions pre- and post-operatively.

Table 1: Demographics of all patients who were sent both versions of TSK-11 questionnaires

	Number (n)	Mean Age(years) +/-SD	Mean BMI +/- SD	Gender		Side		ASA		
				M	F	R	L	1	2	3
Total:	20	66.8 +/- 6.9	30.6 +/- 5.8	8	12	9	11	3	14	3
Hip	8	69.1 +/- 7.9	27.1 +/- 6.0	1	7	3	5	2	6	0
Knee	12	65.2 +/- 6.0	32.8 +/- 4.6	7	5	6	6	1	8	3

BMI: Body mass index; ASA: American Society for Anaesthesiology grade

Table 2: Demographics of patients who completed the TSK-11 questionnaire

	Hip	Knee	Statistical test (p-value)
Total number of patients	37	37	
Age			
Mean	67.2	68.0	2 sample T-test (p=0.3796)
SD	7.1	12.6	
n	36	33	
BMI			
Mean	29.8	32.4	2 sample T-test (p=0.0235)
SD	6.1	5.2	
n	37	37	
Gender			
Male	11	21	Pearson Chi square test (p=0.019)
Female	26	16	
n	37	37	
Side of Operation			
Right	14	16	Pearson Chi square test (p=0.705)
Left	22	21	
n	36	37	
ASA grade			
ASA 1	4	2	Pearson Chi square test (p=0.672)
ASA 2	29	30	
ASA 3+	4	5	

BMI: Body mass index; ASA: American Society for Anaesthesiology grade

In 77.7% of Unsure responses pre-operatively, patients completed the correct version of the TSK-11 by either answering Somewhat Agree (A) or Somewhat Disagree (D) in place of Unsure. In other words, the majority of responses shifted one column to the left or right. In 22.3% of cases, the response was shifted two columns from the middle Unsure column, to Strongly Agree (SA) or Strongly Disagree (SD). There were two patients (JP077, JP078) who accounted for three of the four responses where there was a shift by two columns (see Table 3). If such patients were considered potential outliers, then over 93% of responses shifted by one column only, i.e from Unsure to A or D. The majority of responses were from Unsure to D (75%). In the post-operative group, all of the responses shifted by one column, from Unsure to either A or D. There were also less Unsure responses (6.2%) in the post-operative group compared to the pre-operative group (15.6%).

There were a small minority of responses which changed from the incorrect to the correct version of the TSK-11. In the pre-operative group, 6.9% of responses were altered and in the post-operative group, 3.3% of responses were altered. Table 2 shows how the responses were altered, with four responses involving a shift from Strongly Agree to Strongly Disagree at the pre-operative timepoint. Pre-operatively, in 12 out of the 16 (75%) altered responses, there was a shift from agreement to disagreement and vice versa. Post-operatively, four of the seven (57%) altered responses involved a shift from agreement to disagreement and vice versa. Again, there were two patients who accounted for the majority of these extreme shifts in responses (JP077, JP078).

There were two patients who accounted for a majority of Unsure responses, especially in the pre-operative period (JP065, JP071). In the case of JP071, seven of the 11 questions of the TSK-11 were Unsure. For JP065, six of the 11 questions were Unsure. Together, these two patients accounted for 36% of all Unsure responses pre-operatively. Pre-operatively, questions 2 and 8 had the greatest number of Unsure responses but almost all of the questions had an Unsure response except for question 4. Post-operatively, there was a spread of Unsure responses across all of the questions apart from questions 2,3 and 10.

Table 3: Pattern of Unsure responses to alternative responses and change in response pattern from using incorrect to correct TSK-11

	Pre-Op	Post-Op
No. of cases (n)	21	19
Total no. of responses (n x 11)*	231	209
Total no. of U responses (%)	36 (15.6)	13(6.2)
No. of responses U to A	7	7
No. of responses U to D	21	6
No. of responses U to SA	4	0
No. of responses U to SD	4	0
Total no. of altered responses (%)	16 (6.9)	7 (3.3)
No. of responses SD to D	1	3
No. of responses D to SA	2	1
No. of responses A to SA	1	0
No. of responses SA to D	3	0
No. of responses SA to SD	4	0
No. of responses A to SD	2	2
No. of responses SD to SA	1	0
No. of responses D to SD	2	0
No. of responses D to A	0	1

Unsure (U), Somewhat Agree (A), Somewhat Disagree (D), Strongly agree (SA), Strongly disagree (SD)

*This corresponds to the no of cases multiplied by the 11 questions that make up the TSK-11 questionnaire

A sensitivity analysis was performed (Table 4) which showed that there was only a small difference between the worst and best score compared to the corrected total score for the TSK-11.

Table 4: Sensitivity analysis looking at worst and best score outcomes

	Total score: Worst +/- SD (95%CI) (n=74)	Total score: Best +/- SD (95%CI) (n=74)	Total score: Unsure +/- SD (95%CI) (n=74)	Total score: Corrected +/- SD (Min-Max) (n=20)
Mean TSK-11 score	24.7 +/- 0.8 (23.1 – 26.2)	22.6 +/- 0.7 (21.1 – 24.1)	27.7 +/- 1.1 (25.5-29.8)	24.1 +/- 8.1 (13-44)

The most common number of Unsure responses were one, two or three items in the TSK-11 questionnaire (Table 5). After this, it was rare to see more than 4 items being answered Unsure.

Table 5: Frequency of Unsure responses

No of Unsure responses	Frequency
0	15
1	23
2	11
3	12
4	3
5	5
6	2
7	1
8	2
9	0
10	0
11	0

Table 6 shows the pattern of Unsure responses comparing the whole cohort versus the patients who had completed both the correct and incorrect versions.

Table 6: Number of Unsure responses for each question of the TSK-11 questionnaire

Question	Entire cohort (n=74)	Correct and Incorrect version cohort (n=21)
Q1: I'm afraid I might injure myself if I exercise	21	4
Q2: If I were to try to overcome it, my pain would increase	24	5
Q3: My body is telling me I have something dangerously wrong	9	1
Q4: People aren't taking my medical condition serious enough	4	0
Q5: My accident/problem has put my body at risk for the rest of my life	11	3
Q6: Pain always means I have injured my body	10	3
Q7: Simply being careful that I do not make any unnecessary movements is the safest thing I can do to prevent my pain from worsening	14	4
Q8: I wouldn't have this much pain if there wasn't something potentially dangerous going on in my body	14	6
Q9: Pain lets me know when to stop exercising so that I don't injure myself	18	4
Q10: I can't do all the things normal people do because it's too easy for me to get injured	11	3
Q11: No one should have to exercise when he/she is in pain	16	3

5.4 Discussion

One of the objectives of POSt-pilot was to identify any issues with the use of the questionnaires in the study. The identification of an error in the TSK-11 questionnaire being used helps justify the reason for undertaking a pilot study in the first place. As a result of the pilot study, the use of an incorrect version of the TSK-11 was discovered

and an attempt at trying to mitigate the error was made with the goal of trying to still use the data that had been collected as much as possible. By giving patients both the incorrect and correct TSK-11 versions, data from both versions could be analysed in order to see if there was a pattern in how patients who responded with Unsure in the incorrect version, would subsequently respond in the correct version.

There was a difference in the proportion of patients who responded with Unsure in the preoperative time period compared to the post-operative period, with almost twice as many Unsure responses in the pre-operative period compared to post-operative period. This may reflect the fact that patients were more familiar with the TSK-11 questionnaire a second time around and more able to give a definitive response rather than an Unsure response. The issue of the neutral response in the context of its use in a Likert scale, in this case, Unsure has been a source of extensive debate. Originally, the addition of the neutral response was originally intended to enable respondents to avoid giving a forced false response(undefined author1987 n.d.). In this way, participants could avoid providing a response that was not a true reflection of their beliefs if they were either ignorant or ambivalent to the question or subject matter posed. Studies have shown however, that given a choice of a neutral or unsure response option, there is a significant increase in the number of participants who choose the neutral response even though they may actually have an alternative opinion(Edwards and Smith 2011).

There are three main reasons for why participants may choose to respond in a neutral way. Firstly, people have a tendency to satisfice, in other words, will avoid having to make a conscious effort to interpret a question, digest information and apply their own experiences and opinions in order to choose a non-neutral response on a Likert scale(Krosnick et al. 2002). It is easier to choose a neutral response especially in less motivated participants as it requires less effort on their part(Johns 2005). Secondly, people may choose the neutral response as a result of ambivalence. They may wish to avoid negative feelings associated with conflicted opinions on a particular subject matter. Thirdly, people may choose a neutral response in order to fit in socially and avoid choosing a response which is less socially acceptable or desirable(Krosnick et

al. 2002). Thus, even though the idea of introducing a neutral response in the context of a Likert scale, was to try to encourage participants to avoid providing a false response, the reality is that the neutral response actually may encourage a false response as participants avoid providing a response that truly reflects their opinions for the reasons provided above.

The results showed that the majority of responses for Unsure in the pre-operative period (77.7%) shifted by one column, either to Somewhat Agree or Somewhat Disagree. Only a minority of Unsure responses (22.3%) shifted to the more extreme Strongly Agree and Strongly Disagree. This would seem logical as patients who responded with Unsure are likely to be unsure about how to answer a particular question and if forced to provide an alternative, more definitive response, are more likely to respond one side either way to the middle response of Unsure, in this case, either Somewhat Agree or Somewhat Disagree. The results from this experiment would suggest that participants respond in a wide variety of ways and that it would be impossible to predict how a participant would respond a second time around if they had originally responded in a neutral (Unsure) manner.

Another possible method of attempting to mitigate the error of the presence of a neutral response would be to see if multiple imputation could be used to predict how a participant would respond, using the other items of the TSK-11 questionnaire to see if that might predict a response. If one assumed that U responses were equivalent to missing data, then the principle of using multiple imputation with suitable predictor variables may enable data from participants who had erroneously completed the incorrect TSK-11 questionnaire and responded with an Unsure response, to still include their data for analysis. There has been increasing interest in utilising multiple imputation (MI) techniques to manage missing data in Likert type data. Traditionally, MI methods for dealing with missing data assumes the variables in the model are normally distributed and multivariate. Likert type responses produce data that is typically not normally distributed. Leite and Beretvas(Leite & Beretvas 2010) have shown however that MI can be robust in managing Likert type missing data and

supports the assertion of Schafer who concluded that a MI model could be used successfully for ordered categorical data(Schafer 1997).

Table 7: Recommendations for POSt-Main study

Check the questionnaire and reference and version/date. Ensure that the questionnaire rules are understood and recorded. Ensure the necessary permissions for using the questionnaire are granted.
Send the questionnaire to another member of the research team (or with the trial management group) as a due diligence check
Perform a pilot study with the questionnaires to check for errors or problems with the use of the questionnaires
If an error has occurred, determine precisely what the error is and document this as a study site file entry
Undertake an analysis of the data with the incorrect questionnaire to determine the pattern of incorrect data and determine if the error is likely to be occurring at random (similar to missing data at random pattern)
Try multiple imputation techniques to analyse the data, treating the incorrect data as missing data

5.5 Conclusion

The results of this experiment suggest that there is no clear pattern on how a participant who responded Unsure would have responded if the Unsure response was not a given option. The use of multiple imputation might be a method of predicting a response rate and should be the subject of further research in the future.

Chapter 6 Managing missing data in POSt Pilot

6.1 Introduction

Missing data represents a significant challenge for any research study, no matter how well designed and irrespective of the diligence and care taken by the research team in collecting data. It can adversely influence the ability of a research study to draw definitive conclusions and can lead to selection bias and a loss of sample size. Historically, the impact of missing data and how researchers dealt with this issue has received little attention and as such, has compromised the conclusions reached by many of these clinical studies. Missing data can be defined as “values that are not available and that would be meaningful if they had been observed” (Little et al. 2012).

There are many reasons why missing data occurs(Kaushal 2014). These can be divided into participant factors, data collection process and research study factors(Little et al. 2012; National Research Council 2011; Patel et al. 2015). Participants in any study can be lost to follow up, may withdraw from the study or not answer questionnaires completely(Pedersen et al. 2017). This will inevitably lead to some missing data. The data collection process in the study may be flawed or poorly designed increasing the risk of having missing data. For example, the case report forms may be overly complicated or difficult to complete. Questionnaires may be difficult to read, unclear or ambiguous, leading to difficulties for participants trying to fill the forms accurately and completely(Hardy et al. 2009). Indeed, questionnaires may be lost in the mail system of large organisations. Finally, the research study may face logistical issues and the study team maybe unable to collect the data at the right time or there are issues with consistency between research team members and how they are trained to collect data(Osborne 2012).

Many studies now use patient reported outcome measures (PROMs) as the primary outcome measure. These are usually questionnaires designed and validated in order to provide a patient reported measure on the outcome of an intervention. They consist

of a number of items (questions) which are added together to provide a composite total score. The items are usually ordinal in nature. The Oxford Knee (OKS) and Hip scores (OHS) are examples of multi-item PROMs (see Figure 1).

The Oxford scores are 12-item questionnaires validated as PROMs for use in patients who have undergone a total knee or hip replacement. Most PROMs have scoring rules and missing data rules to enable researchers to calculate the overall composite score. In the example of the Oxford scores, up to two items can be missing yet the composite score can still be calculated using the mean of the total score from the remaining answered items(Murray, Fitzpatrick, Rogers, Pandit, Beard, Carr & Dawson 2007b; Curran et al. 1998). However, if more than two items are missing then the composite score cannot be calculated(Fayers et al. 1998). Other PROMs do not allow for any missing data at all. Missing data in multi-item PROMs can take place at either the item level (in which some items are non-responses) or at the unit level (where all of the items are non-responses) (Rombach et al. 2018).

Missing PROMs data can create hurdles for data analysis, affects how the data is interpreted and reduces the value of the findings(Fayers et al. 1998). High levels of missing PROMS data can cause the following problems:

- It can reduce the power of the study increasing the standard error and the risk of false negative findings (a type 2 error) (174,175).
- Missing data may be related to the measured outcome which can lead to a biased estimate(Fairclough et al. 1998).
- The missing data may not be due to random chance and this again can lead to bias in the analysis(Mercieca-Bebber et al. 2016).
- It can undermine randomisation in a RCT, especially in the setting of an intention to treat analysis, making it less valid, as assumptions about the missing data is required and this is not always verifiable(Bell et al. 2014).

In a review of 71 clinical trials published in the four highest impact medical journals (British Medical Journal, Lancet, Journal of the American Medical Association and

New England Journal of Medicine), 89% reported having some missing data and 18% of the studies had more than 20% of their study participants with missing outcomes(Wood et al. 2004). There was no reporting on how missing data was handled in 66% of clinical trials indexed in PubMed(Chan & Altman 2005).

6.1.1 Types of missing data

There are three main reasons for why data is missing (missingness mechanisms), as first described by Little and Rubin(Little & Rubin 1987).

Missing completely at random (MCAR)

The missingness of the outcome of interest is unrelated to the observed or unobserved patient data, i.e there is no relationship between the missing data and the participant(Bell et al. 2014). For example, the reasons why there was missing OKS data was because there were postal problems and the questionnaire was never completed as a result or the participant never received the OKS because the researcher forgot to send it out(Bell & Fairclough 2014). Another example would be the accidental omission of an answer on a questionnaire as a result of not looking at a double-sided questionnaire form.

Missing at random (MAR)

The missingness is to do with the participant but can be predicted from other information about the person. In other words, the propensity for a data point to be missing is not related to the missing data, but it is related to some of the observed data. For example, in cancer clinical studies, the size of the tumour is often missing but the type of tumour is usually fully recorded. The size of the tumour may be related to the type of tumour thus if the probability that the missing data on the tumour size only depends on the type of primary tumour, then the missingness is considered to be MAR. Another example might be that patients with a high BMI or from a specific ethnic group have a higher incidence of missing data and are more likely to have poorer (lower) OKS(Mercieca-Bebber et al. 2016).

Missing not at random (MNAR)

The missingness (often termed “non-ignorable”) is specifically related to what is missing and the failure to observe a value depends on the value that would have been observed or on other missing values in the data set. The commonest situations where MNAR is encountered are in longitudinal studies in which the missingness is due to participant drop out, death or complications as a result of treatment interventions(Ibrahim et al. 2012). Another example is that patients with poorer health may be less likely to complete a health quality of life questionnaire(Leurent et al. 2018) or patients who are severely affected with pain and dysfunction because of their osteoarthritis may be less likely to complete their OKS fully.

6.1.2 Statistical techniques for handling missing data

Traditionally, there have been a number of ways of managing missing data in PROMs ranging from a complete case analysis to model-based analysis, depending on the assumptions being made for the reasons for the missing data(Pigott 2010). In general, these techniques can be categorised into the following(Rombach et al. 2016):

- Available data (complete case) analysis (CCA)
- Single imputation techniques, replacing the missing observation with a plausible value based on the following premise:
- Last observation carried forward (LOCF), where the missing value is based on previously observed data for that participant
- Mean imputation where the mean of the available observed data is used
- Multiple imputation techniques, using other auxiliary variables that have been observed in order to impute a plausible value to replace the missing value.
- Model based (maximum likelihood) or mixed effect models analysis for longitudinal data

A full review of the different methods of analysing missing data is beyond the remit and scope of this chapter. Instead, I shall concentrate on a few options and explain the reasoning why a particular method was chosen over others. One rather simplistic method is to discount the study participants who have incomplete data and undertake

a complete case analysis (CCA). This is the easiest and commonest method for dealing with missing data but makes the assumption that the missing data is missing MCAR. This may be a reasonable assumption if there are only a few cases with missing data where there is a greater chance of the complete cases representing the population. With greater levels of missing data, this assumption becomes increasingly flawed and can lead to bias. There is evidence that study participants who contribute to the missing data may be different from the participants who submit complete data. Furthermore, by performing a CCA, this can lead to a much smaller sample size with complete data, rendering the conclusions of the data analysis subject to bias, overestimating error and reduces the power of the study(Eekhout et al. 2014).

Another simple method is to use the mean of the observed variable in order to provide a plausible value for the missing data, an example of which is the OHS/OKS, the guidelines of which suggest such a method when there is one or two item level missing data(Murray et al. 2007). This has the major advantage of being simple to use but makes assumptions that the missing data is MAR. It is also limited by the extent of the missing data and if there is more than a certain level of missing data, this technique becomes increasingly subject to bias and most scoring systems such as the Oxford scores only allow for a small amount of missing item data. Furthermore, in the case of the Oxford scores, this technique does not account for unit level missing data.

One method that has gained increasing popularity is the use of multiple imputation (MI), which is an example of a model-based analysis. In 1977, Rubin first introduced the idea of multiple imputation and further elaborated on this in 1987 in his book entitled “Multiple Imputation for Nonresponse in Surveys” (Rubin 2009). Multiple imputation is a powerful simulation based statistical method for handling missing data that uses the distribution of the observed data to estimate a set of plausible values to substitute the missing data. Rubin described a set of rules, “Rubin’s rules”, for combining the individual estimates and standard errors (SE) from each of the *M* imputed datasets into a single overall MI estimate and SE; for example, in determining the significance level for a single combined estimate in hypothesis testing(Rubin

2009). When MI is performed correctly and with the appropriate assumptions, it can generate plausible estimates and standard errors(White et al. 2010).

It consists of two stages. In the first stage (imputation stage), the incomplete dataset is replicated multiple times, with the missing values replaced by plausible values, generated using a range of predictor variables (covariates), based on the correlations between the covariates and the item to be replaced, producing multiple parallel completed data sets(Pigott 2010; Peyre et al. 2011). In the second stage (analysis stage), the appropriate analysis of the imputed data is performed on each of the imputed datasets and the resulting parameter estimates are combined using Rubin's rules(Rubin 2009). Thus, by imputing over and over again, multiple times, MI incorporates missing data uncertainty(Eekhout et al. 2014). The major work involved in multiple imputation involves generating possible values for each missing observation but with modern statistical software packages, MI has become increasingly easier to perform and far less laborious.

In essence, multiple imputation involves the following steps(Allison 2000). In MI, there are two models; the imputation model and the analysis model. The imputation model is used to create imputations in the imputation step. The analysis model is the complete dataset model in which the analysis of the data is undertaken. Using an appropriate model with random variation, missing data is imputed (whereby each missing value or observation within the dataset is substituted with a set of plausible values representing the uncertainty which surrounds the right value to impute). This process is repeated *M* times (usually up to 5 times) to create a *M* number of complete datasets and the required statistical analysis is conducted based on the complete datasets. This process reflects the uncertainty of using these estimated values. Random components are incorporated into the set of the estimated values in order to generate multiple datasets that can then be analysed individually but identically to obtain a set of parameter estimates. The mean of the individual parameter estimates for all of the *M* samples is calculated, in order to produce a single point overall estimate, variance and confidence intervals for the dataset. The standard error is calculated using the following a formula combining (i) the mean of the squared standard errors of the M estimates and (ii) the variance of the *M* parameter estimates across the *M* datasets.

6.1.3 Assumptions for MI

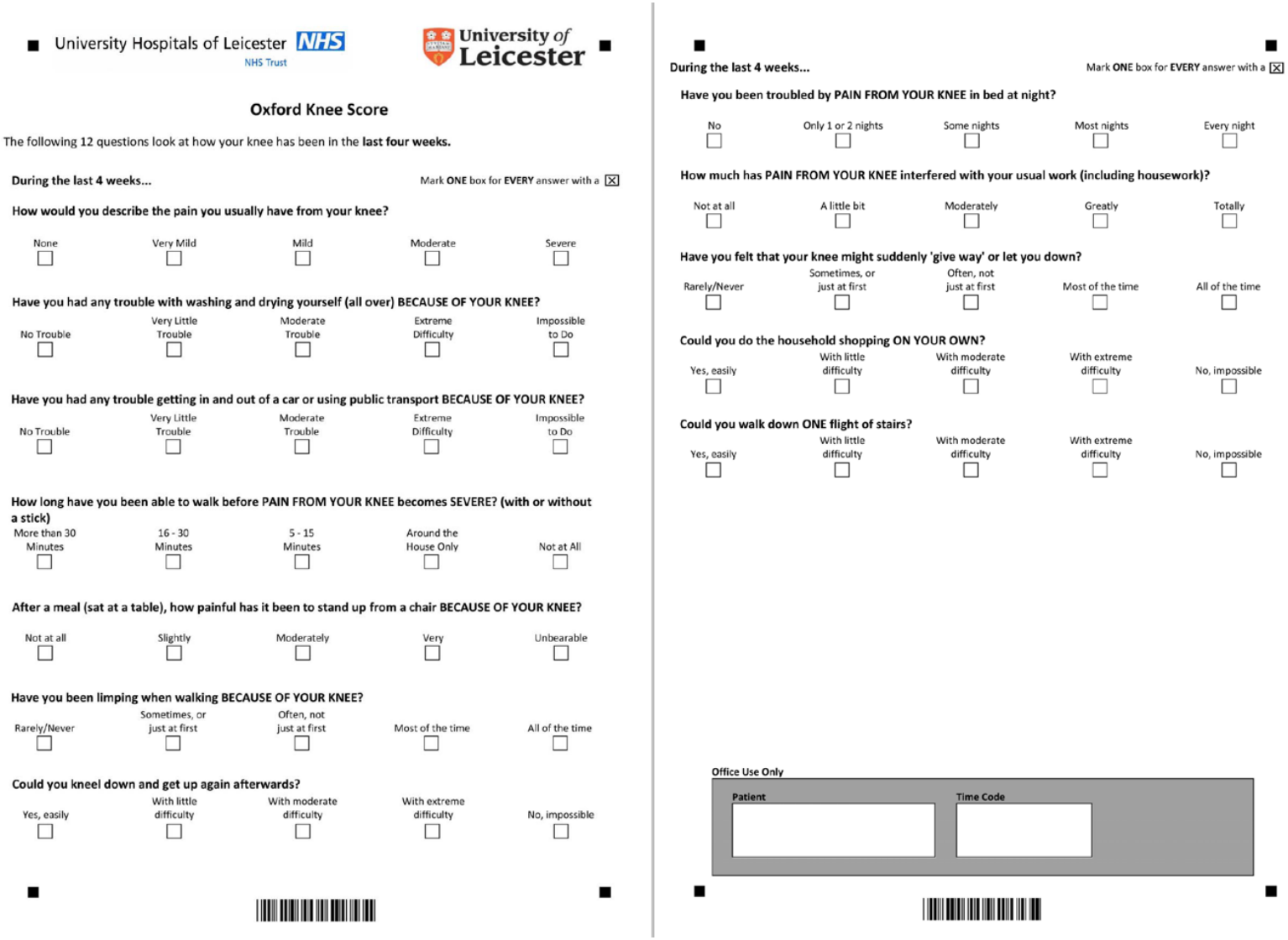
Using MI requires certain assumptions to be made regarding the distribution of the data and pattern of missing data(Sterne et al. 2009). In most cases, it is assumed that the data is being analysed under MAR criteria. In order to avoid the potential for bias, as many predictor variables need to be included in the imputation model, including variables which may predict the missing data as well as variables which influence the process of having missing data, even if these variables may not be included in the final substantive analysis. Sometimes, the data is missing not at random (MNAR) and there is a discrepancy between the analysis results of the complete dataset and that of the imputed dataset. The reason(s) for this need to be identified. This is likely to be as a result of omissions of predictor variables in the imputation model and should be discussed and reported. Another assumption is that the predictor variable data are normally distributed. This can lead to bias if this is not the case. It would therefore be prudent to firstly identify if the predictor variable was normally distributed, and if not, transform such variables to normality before undertaking imputation modelling. This can be a challenge if the data is binary or categorical. It is assumed a categorical variable has the same set of categories and the reference category is the same. Finally, the outcome variable of interest may need to be incorporated into the imputation model as the outcome variable influences the predictor variable and its missing data. For example, in the POSt pilot dataset, the Oxford Knee Score (OKS; outcome variable) may influence Body Mass Index (BMI) and any missing data with the BMI dataset as patients with higher BMIs may have worse OHS but may also be less likely to provide their height and weight as a result of stigmatisation. It would therefore be important to include the OKS as part of the imputation model.

When handling missing PROMs data, MI has been the popular choice and can be utilised at the composite score, subscale and item level(Rombach et al. 2018). At the subscale and item level, MI can help improve the accuracy of the imputation model. There have been many studies that have examined the use of MI in PROMs type questionnaires. A simulation study looking at the EQ-5D-3L health related quality of life questionnaire compared using MI for missing data at the composite and item level in order to assess whether it was better to employ MI at the composite score or item level. In a large dataset (n=1814), the study authors found that the decision to use MI

at the composite score or item level was dependent on the pattern of missing data and the sample size involved. For large sample sizes (n>500), where the pattern of missing data followed a unit non-response, then imputing at the composite score or item level was similar. If the sample size was less than 500, with 10% missing data, again, there was little difference between imputing at the composite score or item level. With higher levels of missing data (20% or more) or if the sample size was small (n< 100), then composite score level imputation was more accurate. If the proportion of item-level missing data increased, then imputation was more accurate if utilised at the item level(Simons et al. 2015). A study in the Netherlands used a dataset of 500 subjects who had completed the Pain Coping Inventory (PCI), a 12-item self-reporting questionnaire, as the basis for a number of simulation experiments comparing different missing data methods including CCA and MI. The authors concluded that in the presence of missing item level data, then MI should be applied at the item level. If there was greater than 10% missing data, then mean imputation methods led to highly biased results(Eekhout et al. 2014).

For this pilot study, we wanted to explore ways of mitigating against missing data, by examining the literature for guidance to improve the design and conduct of the study in order to promote capturing as complete a dataset as possible and also to assess whether using MI was helpful in handling missing data and to recommend an analysis plan on how to handle missing data in the Oxford knee and hip score.

Figure 1: Example of Oxford Knee Score questionnaire used in POSt Pilot



University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Oxford Knee Score

The following 12 questions look at how your knee has been in the **last four weeks.**

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

How would you describe the pain you usually have from your knee?

None ☐ Very Mild ☐ Mild ☐ Moderate ☐ Severe ☐

Have you had any trouble with washing and drying yourself (all over) BECAUSE OF YOUR KNEE?

No Trouble ☐ Very Little Trouble ☐ Moderate Trouble ☐ Extreme Difficulty ☐ Impossible to Do ☐

Have you had any trouble getting in and out of a car or using public transport BECAUSE OF YOUR KNEE?

No Trouble ☐ Very Little Trouble ☐ Moderate Trouble ☐ Extreme Difficulty ☐ Impossible to Do ☐

How long have you been able to walk before PAIN FROM YOUR KNEE becomes SEVERE? (with or without a stick)

More than 30 Minutes ☐ 16 - 30 Minutes ☐ 5 - 15 Minutes ☐ Around the House Only ☐ Not at All ☐

After a meal (sat at a table), how painful has it been to stand up from a chair BECAUSE OF YOUR KNEE?

Not at all ☐ Slightly ☐ Moderately ☐ Very ☐ Unbearable ☐

Have you been limping when walking BECAUSE OF YOUR KNEE?

Rarely/Never ☐ Sometimes, or just at first ☐ Often, not just at first ☐ Most of the time ☐ All of the time ☐

Could you kneel down and get up again afterwards?

Yes, easily ☐ With little difficulty ☐ With moderate difficulty ☐ With extreme difficulty ☐ No, impossible ☐

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

Have you been troubled by PAIN FROM YOUR KNEE in bed at night?

No ☐ Only 1 or 2 nights ☐ Some nights ☐ Most nights ☐ Every night ☐

How much has PAIN FROM YOUR KNEE interfered with your usual work (including housework)?

Not at all ☐ A little bit ☐ Moderately ☐ Greatly ☐ Totally ☐

Have you felt that your knee might suddenly 'give way' or let you down?

Rarely/Never ☐ Sometimes, or just at first ☐ Often, not just at first ☐ Most of the time ☐ All of the time ☐

Could you do the household shopping ON YOUR OWN?

Yes, easily ☐ With little difficulty ☐ With moderate difficulty ☐ With extreme difficulty ☐ No, impossible ☐

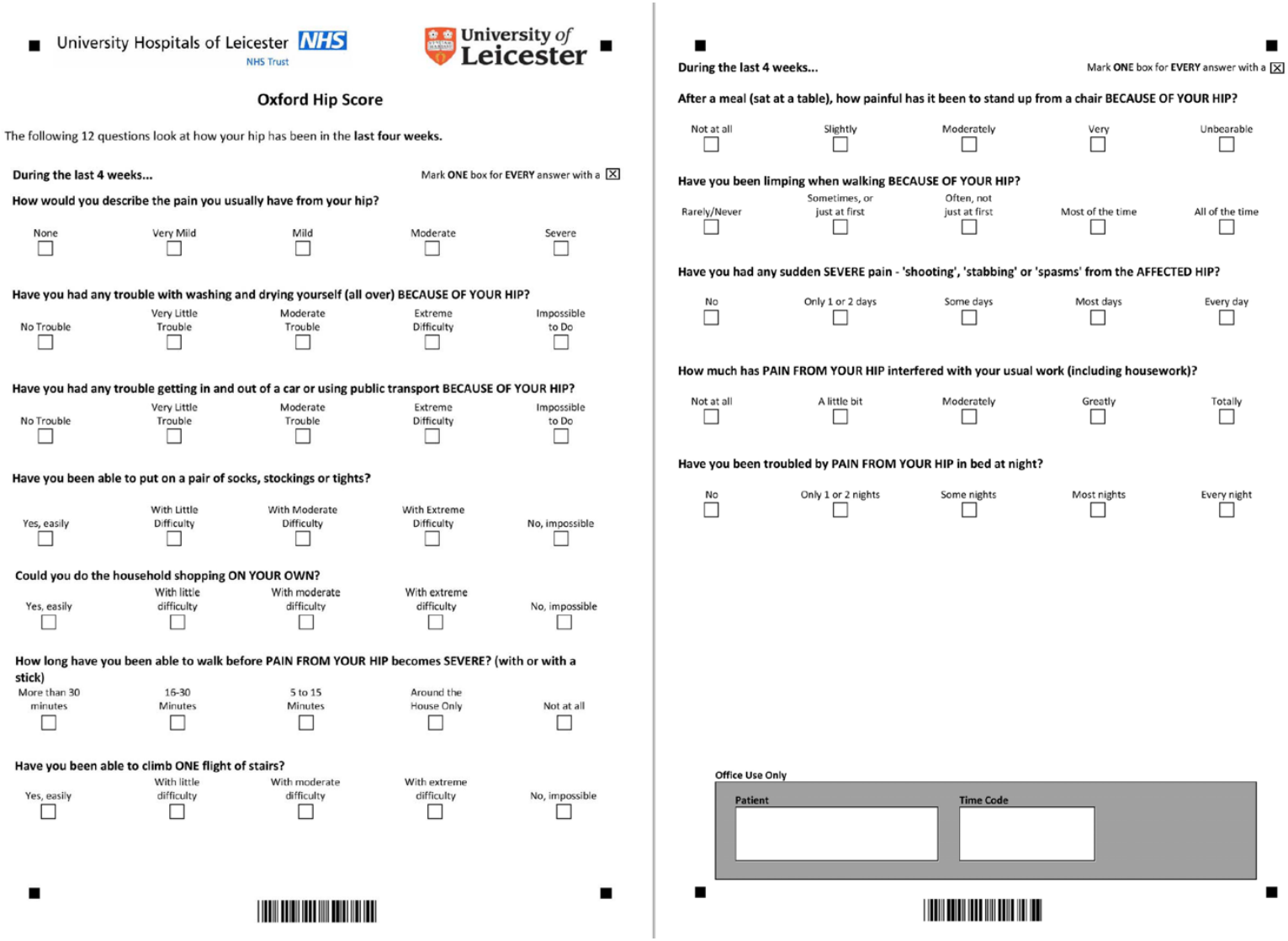
Could you walk down ONE flight of stairs?

Yes, easily ☐ With little difficulty ☐ With moderate difficulty ☐ With extreme difficulty ☐ No, impossible ☐

Office Use Only

Patient **Time Code**

Figure 2: Example of Oxford Hip Score questionnaire used in POSt Pilot



University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Oxford Hip Score

The following 12 questions look at how your hip has been in the **last four weeks.**

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

How would you describe the pain you usually have from your hip?

None ☐ | Very Mild ☐ | Mild ☐ | Moderate ☐ | Severe ☐

Have you had any trouble with washing and drying yourself (all over) BECAUSE OF YOUR HIP?

No Trouble ☐ | Very Little Trouble ☐ | Moderate Trouble ☐ | Extreme Difficulty ☐ | Impossible to Do ☐

Have you had any trouble getting in and out of a car or using public transport BECAUSE OF YOUR HIP?

No Trouble ☐ | Very Little Trouble ☐ | Moderate Trouble ☐ | Extreme Difficulty ☐ | Impossible to Do ☐

Have you been able to put on a pair of socks, stockings or tights?

Yes, easily ☐ | With Little Difficulty ☐ | With Moderate Difficulty ☐ | With Extreme Difficulty ☐ | No, impossible ☐

Could you do the household shopping ON YOUR OWN?

Yes, easily ☐ | With little difficulty ☐ | With moderate difficulty ☐ | With extreme difficulty ☐ | No, impossible ☐

How long have you been able to walk before PAIN FROM YOUR HIP becomes SEVERE? (with or with a stick)

More than 30 minutes ☐ | 16-30 Minutes ☐ | 5 to 15 Minutes ☐ | Around the House Only ☐ | Not at all ☐

Have you been able to climb ONE flight of stairs?

Yes, easily ☐ | With little difficulty ☐ | With moderate difficulty ☐ | With extreme difficulty ☐ | No, impossible ☐

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

After a meal (sat at a table), how painful has it been to stand up from a chair BECAUSE OF YOUR HIP?

Not at all ☐ | Slightly ☐ | Moderately ☐ | Very ☐ | Unbearable ☐

Have you been limping when walking BECAUSE OF YOUR HIP?

Rarely/Never ☐ | Sometimes, or just at first ☐ | Often, not just at first ☐ | Most of the time ☐ | All of the time ☐

Have you had any sudden SEVERE pain - 'shooting', 'stabbing' or 'spasms' from the AFFECTED HIP?

No ☐ | Only 1 or 2 days ☐ | Some days ☐ | Most days ☐ | Every day ☐

How much has PAIN FROM YOUR HIP interfered with your usual work (including housework)?

Not at all ☐ | A little bit ☐ | Moderately ☐ | Greatly ☐ | Totally ☐

Have you been troubled by PAIN FROM YOUR HIP in bed at night?

No ☐ | Only 1 or 2 nights ☐ | Some nights ☐ | Most nights ☐ | Every night ☐

Office Use Only

Patient | **Time Code**

6.2 Methods

6.2.1 Body Mass Index and statistical analysis plan for missing data

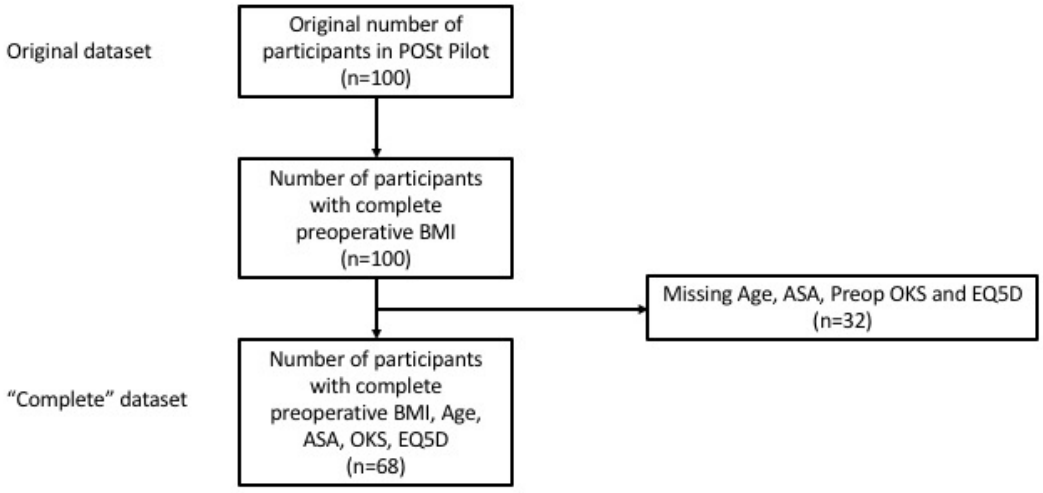
Body mass index (BMI) was chosen as the variable to test in different simulation models comparing a complete dataset (with complete data for BMI and the predictor variables) against a dataset where BMI had missing data introduced in increasing amount and with a dataset where BMI had imputed data. Three groups of data were created (Table 1) with a complete dataset, missing dataset with increasing levels of missing data and finally an imputed dataset which used multiple imputation (MI) to impute missing BMI observations. The complete dataset had 68 cases with a full set of observations for all these variables, available for analysis. Using a Stata (StataCorp. 2013) coding program, random missing observations in the BMI variable were artificially created initially with 10% of observations missing, then 20%, 30%,40% and finally 50% missing data (see Table 1).

Table 1: Description of BMI groups created for the simulation study (n=68)

Group	**Description**	**% BMI data missing**	**Number of observations missing (n)**
Group 1	Complete BMI data “Complete”	0	0
Group 2	Missing BMI data “Missing”		
2a		10	7
2b		20	14
2c		30	20
2d		40	27
2e		50	34
Group 3	Multiple imputation data “Imputed”	NA	NA

NA: Not applicable; BMI: Body Mass Index

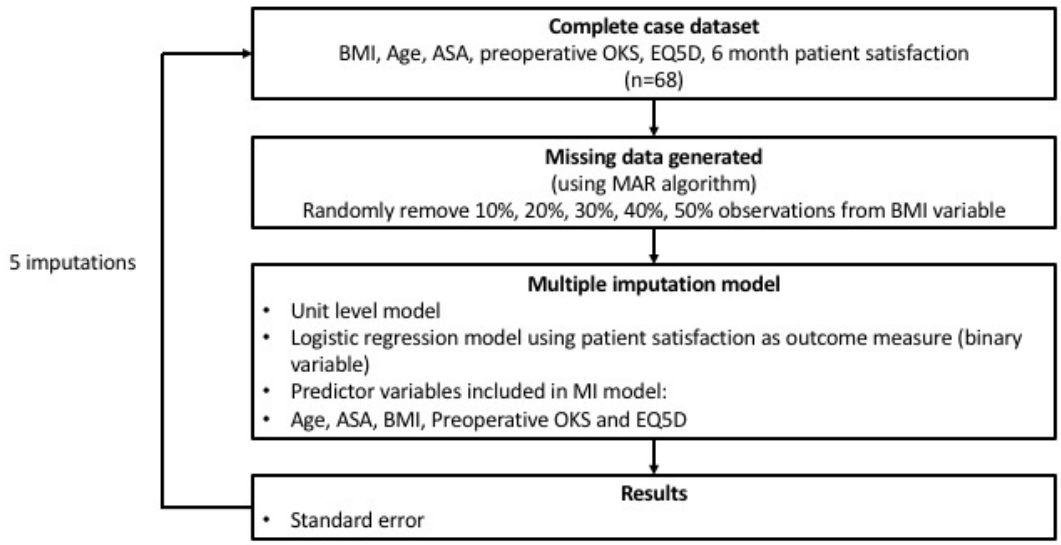
Figure 3: Flowchart of simulation study for BMI showing datasets



BMI: Body Mass Index; ASA: American Society for Anaesthesiology grade; OKS: Oxford Knee Score; EQ5D: European Quality of Life health quality questionnaire

Ordered logistic regression analysis was performed for Groups 1 and 2 to with patient satisfaction at six months as the primary outcome, which is a binary variable: either satisfied or dissatisfied. The covariates used in the regression model were: Age (continuous data), BMI (continuous data), ASA grade (ordinal categorical), preoperative OKS (continuous data) and preoperative EQ5D (continuous data). Multiple imputation was then performed (Figure 4) using 5 imputations to provide imputed observations for the missing BMI data. The same ordered logistic regression analysis was then repeated using the imputed dataset. The standard errors (SE) from the regression analysis from each group analysis was then compared between Groups 1 to 3.

Figure 4: A flowchart outlining the simulation study for the MI model for missing BMI data



BMI: Body Mass Index; ASA: American Society for Anaesthesiology grade; OKS: Oxford Knee Score; EQ5D: European Quality of Life health quality questionnaire

It would be expected that Group 1 is the most "precise" analysis as it uses complete data. Group 2 would be the least precise dataset and it would be expected that the standard error (SE) would increase with increasing levels of missing data in the regression analysis. Group 3 would be expected to have SE in between that of the "complete" Group 1 and "missing" Group 2 datasets.

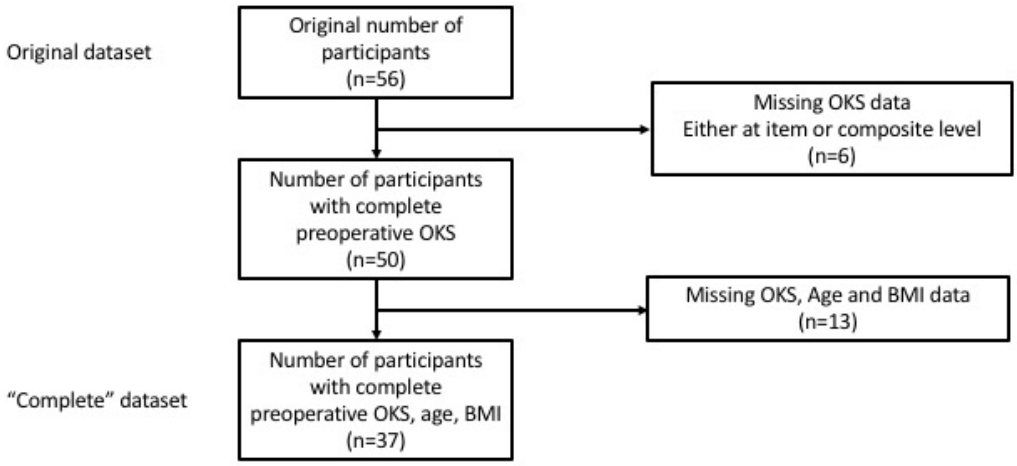
6.2.2 Oxford Knee Score and statistical analysis plan for missing data

Data from POSt Pilot was analysed and simulation models were performed to explore methods of dealing with missing data from the Oxford Knee Score questionnaire. For the pilot study, a decision was made to focus on just the preoperative time period and to assess the pattern of missing data in knee patients who had completed the OKS. The assumption was that the same imputation process could be undertaken with hip

patients, but it would not be possible to combine the hip and knee patient dataset as the OKS and OHS, whilst similar, contain different item level questions. The knee group was chosen as the basis for the simulation model as the sample size with preoperative OKS data was larger (n=56) compared to the hip group (n=39). The data on the preoperative OKS was then used as the basis for a simulation exercise in which participants and item level responses were randomly deleted and then imputation performed and compared with the actual complete dataset. The methods section below details the process of the simulation exercise performed.

There were 50 participants (out of the original 56 participants) who had completed the preoperative OKS. There were however some missing data in the Age and BMI variables so the final number of participants with a complete set of data for OKS, Age and BMI was 37 (see Figure 3). It was this dataset that was termed “Complete”. Each component item is scored from 0 to 4 with 0 being the worst score and 4 being the best. A final pre-operative total OKS was then calculated (0-48, with 0 being the worst and 48 being the best score) for each study participant.

Figure 5: Flowchart of OKS simulation study showing sample size of datasets



A number of simulation models were than performed to explore what would happen if missing data were randomly introduced into the “Complete” dataset. One participant (out of the 37) was chosen at random to have 1 of the 12 component questions that make up the 12-item OKS, to be randomly deleted, thereby simulating a scenario in which 1 of the 12 items was missing in 1 participant. A random number generator (www.random.org; Dr Mads Haahr; School of Computer Science and Statistics at Trinity College, Dublin, Ireland) was used to select the study participant and items to have the data deleted for each of the simulation scenarios.

The Oxford group have recommended that in the event that one or two items (of the 12) are missing, then it would be reasonable to calculate the rest of the items and find the mean which would then be used to fill the missing data of the one or two items, then calculate the total OKS(Murray et al. 2007). These rules are termed “Oxford rules” in this thesis. If three or more items have missing data, then the total OKS cannot be calculated.

In this series of simulation models, for one randomly selected participant, one item, then two items then three items were randomly selected to have their data deleted, representing an increasing number of missing items in the OKS.

Two methods were chosen to manage this missing data. The first method involved using the Oxford rules for dealing with missing data in one or two items. This was also done in the simulation model where 3 items were missing in order to provide a comparison with the 2nd method, which involved using multiple imputation by chained equations (MICE), to impute the missing data.

Table 2: Summary of simulation models used in the analysis, for pre-operative OKS in knee patients.

Simulation model	Number of participants with complete data	No. of participants with missing data	No. of OKS items missing data	OKS rules allowed	Multiple Imputation
1	36	1	1	Yes	Yes
2	36	1	2	Yes	Yes
3	36	1	3	No	Yes

Multiple imputation analysis plan

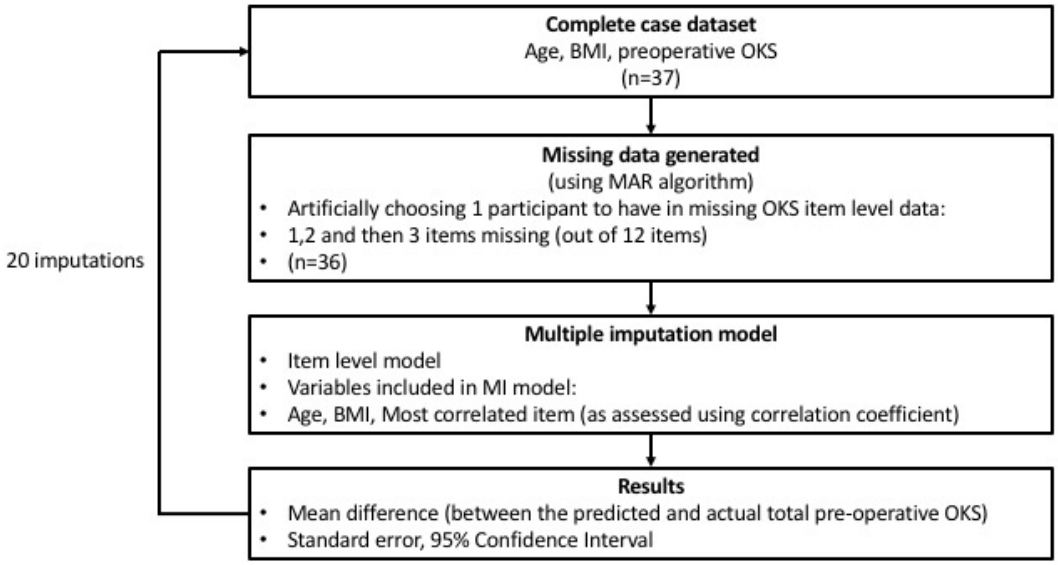
As the sample size for the simulation models was small (n=50), it was not possible to include all of the predictor variables that would ideally have been included in the imputation model. This would include the remaining question items with data. For example, if Q1, Q4 and Q5 for a study participant had missing data, the other 9 questions that make up the 12-item OKS could be used as predictor variables for the imputation model. A correlation matrix for each of the 12 questions (defined as "item" in this thesis) in the OKS was produced using the Spearman correlation method and a correlation coefficient was produced which identified which item was the most strongly correlated with a particular OKS item. This was performed for all 12 items and the results are shown in Table 3.

There were situations where due to the random number predictor, the predictor variables included in the MI model had missing data. In this situation, there was a problem with non-convergence in the MI model. The method chosen to manage this was to use the 2nd most correlated item which had complete data, in the MI model.

In the event that one of the predictor variable items was missing but could be imputed, then the imputed figure was used for that missing item. The imputed item was then

used a predictor variable for the main MI model. In effect, this is akin to using MICE to impute the missing data in the model but only done once whereas the MICE method would repeat this process over and over again giving a far more precise imputation output. As this was a pilot study to explore using MI, 20 imputations were performed as the sample size was small. The predictor variables used in the MI model were: Age, Body Mass Index and one item (of the OKS) that appeared to most correlate with the missing item(s), as discussed above. In order to calculate a linear score for each observation, the weights for this score was calculated using the regression coefficients of an ordered logistic regression model predicting the occurrence of the missing data pattern, using the covariates that are predictive of missing data. Figure 6 summarises the MI simulation model.

Figure 6: A flowchart outlining the simulation study for the MI model for missing OKS data



Data was analysed using Stata Version 13.1 (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP). Multiple imputation was undertaken using the STATA software.

Table 3: A summary of which pre-operative OKS item is most strongly correlated to one another, using the Spearman correlation coefficient.

OKS Question number	**Component item of OKS (description)**	**Strongest correlated OKS question number**	**Component item of OKS (description)**	**Spearman correlation coefficient**
1	How would you describe the pain you usually have from your knee?	4	For how long have you been able to walk before pain from your knee becomes severe? (with or without a stick)	0.6553
2	Have you had any trouble with washing and drying yourself (all over) because of your knee?	12	**Could** you walk down one flight of stairs?	0.5308
3	Have you had any trouble getting in and out of a car or using public transport because of your knee? (whichever you would tend to use)	5	After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your knee?	0.7163
4	For how long have you been able to walk before pain from your knee becomes severe? (with or without a stick)	12	**Could** you walk down one flight of stairs?	0.6905
5	After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your knee?	7	**Could** you kneel down and get up again afterwards?	0.7323
6	Have you been limping when walking, because of your knee?	10	Have you felt that your knee might suddenly 'give way' or let you down?	0.5320
7	**Could** you kneel down and get up again afterwards?	12	**Could** you walk down one flight of stairs?	0.7522

8	Have you been troubled by pain from your knee in bed at night?	5	After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your knee?	0.5889
9	How much has pain from your knee interfered with your usual work (including housework)?	7	**Could** you kneel down and get up again afterwards?	0.7119
10	Have you felt that your knee might suddenly 'give way' or let you down?	6	Have you been limping when walking, because of your knee?	0.5320
11	**Could** you do the household shopping on your own?	12	**Could** you walk down one flight of stairs?	0.8201
12	**Could** you walk down one flight of stairs?	11	How much has pain from your knee interfered with your usual work (including housework)?	0.8201

6.3 Results

6.3.1 Body Mass Index

There were 68 patients in the sample size. The standard errors (SE) for Group 1 was the lowest, as expected, given that this analysed data from a complete dataset. The SE for the missing dataset (Group 2) was the highest as more data was missing (from 10% missing BMI data to 50% missing), again as expected. The imputation model provided SE which were similar, if slightly worse compared the SE seen with 10% and 20% missing data (Table 4). It should be noted that there were some instances in the regression model where non-convergence of the regression model took place. This is

likely to be related to the small sample size and using five covariates as predictor variables in the model.

Table 4: Standard errors comparing Groups 1 to 3

% Missing BMI data	**Standard Error (SE)**		
	Group 1 "Complete"	**Group 2 "Missing"**	**Group 3 "Imputed"**
10	0.0893379	0.09279524	0.10434372
20	0.0893379	0.0931979*	0.12054842
30	0.0893379	0.12969935*	0.126095117
40	0.0893379	0.202806725*	0.16497228
50	0.0893379	0.347831067*	0.1987466

SE: Standard error; BMI: Body Mass Index; MI: Multiple imputation

*Some non-convergence results

As the level of missing data approached over 30%, then the SE for the imputation method was less than that seen in Group 2 with the missing data.

6.3.2 Prevalence of missing Oxford Knee Score and Hip score data

For the knee group, of the 56 participants who had completed preoperative, three month and six-month post-operative OKS, 50 had a completed the OKS fully. Table 5 details which items were missing for the OKS dataset at the different timepoints. There is a spread of items missing from Question 1 to 12 in the OKS with no specific item missing disproportionately more than the other items.

Table 5: Prevalence of missing data (n=56) for OKS for individual items (Q1 to Q12) at different timepoints

Item	**Pre-Operative**		**Three months**		**Six months**	
	No. missing	% missing	No. missing	% missing	No. missing	% missing
Q1	5	8.9	11	19.6	14	25.0
Q2	5	8.9	11	19.6	12	21.4
Q3	5	8.9	11	19.6	12	21.4
Q4	5	8.9	11	19.6	12	21.4
Q5	5	8.9	11	19.6	12	21.4
Q6	5	8.9	11	19.6	14	25.0
Q7	5	8.9	14	25.0	14	25.0
Q8	5	8.9	10	17.9	13	23.2
Q9	5	8.9	10	17.9	13	23.2
Q10	6	10.7	10	17.9	13	23.2
Q11	6	10.7	11	19.6	13	23.2
Q12	6	10.7	10	17.9	13	23.2

In the hip group, the prevalence of missing data was slightly different compared to the knee patients. Table 6 shows the prevalence of missing data, and at the preoperative timepoint, the missing items were mainly question 6 to question 12. This was different at the three and six month follow up where there were all items had missing data, and as with the knee group, there was no one item that had a higher than expected proportion of missing data.

Table 6: Prevalence of missing data (n=39) for OHS for individual items (Q1 to Q12) at different timepoints

Item	**Pre-Operative**		**Three months**		**Six months**	
	No. missing	% missing	No. missing	% missing	No. missing	% missing
Q1	0	0	5	12.8	7	17.9
Q2	0	0	4	10.3	7	17.9
Q3	0	0	4	10.3	7	17.9
Q4	0	0	4	10.3	8	20.5
Q5	0	0	4	10.3	10	25.6
Q6	1	2.6	5	12.8	7	17.9
Q7	0	0	4	10.3	8	20.5
Q8	1	2.6	6	15.4	9	23.1
Q9	2	5.1	6	15.4	8	20.5
Q10	1	2.6	6	15.4	8	20.5
Q11	1	2.6	6	15.4	7	17.9
Q12	2	5.1	6	15.4	7	17.9

6.3.3 Pattern of missing data in the OKS

Approximately 10.7% of the OKS data in the preoperative timepoint was missing (Table 7). The majority of missing data occurred when there was unit level missing data (8.9%), i.e all of the items were missing and so a composite total OKS was not possible. This was the same pattern at all timepoints, but with increasing levels of missing data overall. At six months, 28.6% of the OKS data was missing.

Table 7: Pattern of which items of the OKS are missing at different timepoints

Timepoint	**Missing for <u>which</u> items**	**Frequency**	**Percentage (%)**	**Cumulative percentage (%)**
Pre-operative	1,2,3,4,5,6,7,8,9,10,11,12	5	8.93	8.93
	10,11,12	1	1.79	10.71
	No missing data	50	89.29	100.0
	Total	56	100.0	
Three months	1,2,3,4,5,6,7,8,9,10,11,12	10	17.86	17.86
	1,2,3,4,5,6,7	1	1.79	19.64
	7	3	5.36	25.00
	11	1	1.79	26.79
	No missing data	41	73.21	100.0
	Total	56	100.0	
Six months	1,2,3,4,5,6,7,8,9,10,11,12	12	21.43	21.43
	1,6,7	1	1.79	23.21
	1	1	1.79	25.00
	6	1	1.79	26.79
	7,8,9,10,11,12	1	1.79	28.57
	No missing data	40	71.43	100.0
	Total	56	100.0	

In a few cases, six or seven items (out of the 12) were missing. The pattern would suggest that the participant either did not turn the page over and forgot to complete the 1st half or latter half of the 12-item questionnaire and that this was likely to be a MCAR event. Table 8 demonstrates that either participants failed to complete the entire OKS or more rarely only missed out between one and three items of the OKS.

Table 8: Pattern of how many items of the OKS are missing at different timepoints

Timepoint	Missing for how many items?	Frequency	Percentage (%)
Preop	0	50	89.29
	3	1	1.79
	12	5	8.93
	Total	56	100.0
Three months	0	41	73.21
	1	4	7.14
	7	1	1.79
	12	10	17.86
	Total	56	100.0
Six months	0	40	71.43
	1	2	3.57
	3	1	1.79
	6	1	1.79
	12	12	21.43
	Total	56	100.0

6.3.4 Pattern of missing data in the OHS

In the hip group, the pattern of missing data was slightly different to that of the knee patients (Table 9). In the preoperative timepoint, there was less missing data overall (7.7%) and no unit level missing data. At the three and six month timepoints, there was increasingly more unit level missing data (28.2% at three months and 30.8% at six months), similar to that seen in the knee group at the same timepoints (Table 10).

Table 9: Pattern of which items of the OHS are missing at different timepoints

Timepoint	**Missing for <u>which</u> items**	**Frequency**	**Percentage (%)**	**Cumulative percentage (%)**
Pre-operative	6,9	1	2.56	2.56
	8,9,10,11,12	1	2.56	5.13
	12	1	2.56	7.69
	No missing data	36	92.31	100.0
	Total	39	100	
Three months	1,2,3,4,5,6,7,8,9,10,11,12	2	5.13	5.13
	1,2,3,4,5,6,7,8,9,10,11	2	5.13	10.26
	1	1	2.56	12.82
	6	1	2.56	15.38
	8,9,10,11,12	1	2.56	17.95
	8,9,10,11	1	2.56	20.51
	12	3	7.69	28.21
	No missing data	28	71.79	100.0
	Total	39	100.0	
Six months	1,2,3,4,5,6,7,8,9,10,11,12	7	17.95	17.95
	4,5,7	1	2.56	20.51
	5,8	1	2.56	23.08
	5	1	2.56	25.64
	8,10	1	2.56	28.21
	9	1	69.23	30.77
	No missing data	27	100.0	100.0
	Total	39		

Table 10: Pattern of how many items of the OHS are missing at different timepoints

Timepoint	Missing for how many items?	Frequency	Percentage (%)
Preop	0	36	92.31
	1	1	2.56
	2	1	2.56
	5	1	2.56
	Total	39	100.0
Three months	0	28	71.79
	1	5	12.82
	4	1	2.56
	5	1	2.56
	11	2	5.13
	12	2	5.13
	Total	39	100.0
Six months	0	27	69.23
	1	2	5.13
	2	2	5.13
	3	1	2.56
	12	7	17.95
	Total	39	100.0

6.3.5 Feasibility of the MI approach

In order to examine whether MI was feasible, the simulation model focused on the knee dataset at the preoperative timepoint that was complete. Firstly, in order to determine which item would be included in the MI model as a covariate, a correlation analysis was undertaken to see which item of the OKS had the strongest correlation with other items of the OKS. Table 11 shows this in detail, where the green labelled item had the strongest correlation, the yellow labelled item had the 2nd strongest correlation and the red labelled item the 3rd strongest correlation. For the pilot, only

the most strongly correlated item was used in the MI model as the sample size to carry out a full MI model was too small (Table 12). Furthermore, as the sample size was small, if more than one item was missing, the risk of non-convergence of the analysis increased. It was therefore decided to just impute using one item as a covariate, together with age and body mass index as the other two covariates.

Table 11: Shows the Spearman coefficient for each item in the pre-operative OKS relative to the other items.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Q1	1.00 00	0.43 61	0.51 04	0.65 53	0.53 12	0.43 22	0.45 40	0.47 06	0.58 55	0.47 22	0.50 06	0.50 65
Q2	0.43 61	1.00 00	0.43 96	0.49 61	0.46 81	0.21 84	0.41 20	0.51 66	0.33 67	0.25 94	0.43 02	0.53 08
Q3	0.51 04	0.43 96	1.00 00	0.59 07	0.71 63	0.42 54	0.63 33	0.31 47	0.68 78	0.43 48	0.68 55	0.60 80
Q4	0.65 53	0.49 61	0.59 07	1.00 00	0.54 38	0.29 70	0.63 60	0.33 28	0.66 20	0.41 36	0.66 21	0.69 05
Q5	0.53 12	0.46 81	0.71 63	0.54 38	1.00 00	0.39 19	0.73 23	0.58 89	0.60 08	0.34 63	0.56 39	0.53 45
Q6	0.43 22	0.21 84	0.42 54	0.29 70	0.39 19	1.00 00	0.32 32	0.04 93	0.30 62	0.53 20	0.27 97	0.30 22
Q7	0.45 40	0.41 20	0.63 33	0.63 60	0.73 23	0.32 32	1.00 00	0.40 51	0.71 19	0.39 84	0.66 61	0.75 22
Q8	0.47 06	0.51 66	0.31 47	0.33 28	0.58 89	0.04 93	0.40 51	1.00 00	0.40 49	0.05 49	0.42 63	0.30 25
Q9	0.58 55	0.33 67	0.68 78	0.66 20	0.60 08	0.30 62	0.71 19	0.40 49	1.00 00	0.51 31	0.42 63	0.64 14
Q10	0.47 22	0.25 94	0.43 48	0.41 36	0.34 63	0.53 20	0.39 84	0.05 49	0.51 31	1.00 00	0.66 96	0.51 22
Q11	0.50 06	0.43 02	0.68 55	0.66 21	0.56 39	0.27 97	0.66 61	0.42 63	0.66 96	0.45 14	1.00 00	0.82 01
Q12	0.50 65	0.53 08	0.60 80	0.69 05	0.53 45	0.30 22	0.75 22	0.30 25	0.64 14	0.51 22	0.82 01	1.00 00

Table 12: Summary of the three items which are the most strongly correlated with each question in the pre-operative OKS

OKS item	The OKS item that correlates strongest	The OKS item that correlates 2nd strongest	The OKS item that correlates 3rd strongest
1	4	9	5
2	12	8	4
3	5	9	11
4	12	11	9
5	7	3	9
6	10	1	3
7	12	5	9
8	5	2	1
9	7	3	11
10	6	9	12
11	12	3	10
12	11	7	4

Convergence failures were observed for all imputation models when more than one item was included as a predictor variable in the MI model. With a larger sample size, we would aim to include all of the items which had a response in the MI model. We were unable to include any more than three variables in the model because of difficulties with convergence as the sample size was too small. Considerations on how to manage this issue are discussed later in the Discussion.

The mean difference in the composite total OKS was greater when the Oxford rules were followed compared to the MI model, when between one and three items were missing, in a dataset where one participant had missing data (Table 13). The Oxford rules for one or two missing item data provided an estimate of the composite total OKS that was similar to the composite score if there was no missing data with similar standard errors. Similarly, the MI model also provided reasonably close estimates of

the composite score. There was a trend showing that as more items were missing, then the MI model became more accurate compared to the Oxford rules.

Table 13: Shows the mean difference between the predicted and actual total pre-operative OKS when using the OKS rules and multiple imputation modelling methods, with increasing number of missing items from the OKS.

Number of missing items	**Mean difference in total OKS using OKS rules*** **(SE in brackets)**	**95% CI**	**Mean difference in total OKS using MI** **(SE in brackets)**	**95% CI**
1	0.787 (SE 0.144)	0.461 - 1.112	0.450 (SE 0.172)	0.061 - 0.839
2	1.480 (SE 0.262)	0.887 - 2.073	0.925 (SE 0.208)	0.453 - 1.397
3	1.532 (SE 0.319)**	0.810 - 2.254	1.000 (SE 0.365)	0.174 - 1.826

OKS (Oxford knee score); SE (Standard Error); CI (95% Confidence Interval); MI (Multiple imputation)

*Rules for managing missing data in the OKS: Missing values/notes for analysis. If after repeated attempts to obtain complete data from an individual, only one or two questions have been left unanswered, it is reasonable to enter the mean value representing all of their other responses, to fill the gaps.

An alternative computerised method of imputing values has been reported by Jenkinson et al (2006). If more than two questions are unanswered, it is recommended that an overall score should not be calculated. If patients indicate two answers for one question it is recommended that the convention of using the worst (most severe) response is adopted.

** This is a theoretical score if the OKS rules were adopted. In practice, the Oxford team have recommended that if 3 or more component questions are missing, then an overall score should not be calculated and that data is in effect lost.

6.4 Discussion

The National Research Council (NRC) of the National Academies in the USA set up a specific panel on handling missing data in clinical trials in order to provide guidance and recommendations on how best to manage this problem. The panel concluded that

there were two critical elements which needed to be focused upon: (1) design and conduct of clinical studies to minimise the amount of missing data and (2) analysis that made use of all available information on all participants based on an understanding the assumptions about the nature of the missing data(National Research Council 2011). Whilst there are a number of statistical analysis methods for helping to manage missing data, the assumption that such methods can fully compensate for missing data is unjustified and the design and conduct of a clinical study to limit the likelihood of missing data is critically important(Little et al. 2012). The NRC stated that *"investigators, sponsors, and regulators should design clinical trials consistent with the goal of maximizing the number of participants who are maintained on the protocol-specified intervention until the outcome data are collected."* (National Research Council 2011)

Missing data in variables such as BMI is a common occurrence in many studies. The best way to manage such missing data is to endeavour to collect the data as much as possible from different sources (clinical notes, other hospital databases, theatre records etc). With POSt Pilot, where BMI data was missing from the case report forms completed by the research team, then the original clinical and theatre notes were requested, and the BMI was also checked against other databases collected for that particular patient. In this way, the BMI completeness rate was 100% for POSt Pilot. This allowed simulation studies to be conducted to analyse how accurate multiple imputation methods would be if there was some missing data in BMI or another similar variable for the main POSt study. The results from this experiment suggests that MI is useful where high levels of missing data in the BMI is present. Where only 10 to 20% of BMI data was missing, then MI methods may not make such a big difference. It should be noted that because this was a pilot study with limited sample size, non-convergence of the regression model was an issue and for the POSt main study, one would hope that with a much larger samples size (n>400), then non convergence with the number of covariates used in the regression model would not occur.

The number of imputations required remains a matter for debate(Royston & White 2011a; White et al. 2011). Traditional textbooks on imputation suggest that as few as

3-5 imputations (m=3 or 5) are required(Schafer 1997; Azur et al. 2011; Bouhlila & Sellaouti 2013). In the BMI dataset, a decision to use five imputations was made based on this argument(Schafer 1997; White et al. 2011). White et al. suggested that a greater number of imputations are required for the model to be more accurate and that a general rule of thumb is that the percentage of missing data equates to the number of imputations required in the MI model(White et al. 2011). For example, if 10% of the data is missing, then 10 imputations (m=10) would be reasonable. Going forward, one of the outcomes of this pilot analysis is that using this rule of thumb would be a sensible option, so that if 10% of the data is missing, then 10 imputations would be performed and if 20% of the data was missing, then 20 imputations would be performed and so on.

The prevalence of missing data in the OKS at the preoperative timepoint was 10.71%, 26.79% at three months and 28.57% at six months post-operatively. For the hip patients, the prevalence of missing OHS data was 7.69%, 28.21% and 30.77% at the preoperative, three month and six-month post-operative time period respectively. In a large prospective study involving 856 patients who completed an OKS, 19% of the questionnaires were incomplete(Whitehouse et al. 2005). In a registry based study looking at 91936 patients who were registered in the National Joint Registry having had a primary knee replacement, 27.3% had incomplete or missing OKS data had had to be excluded from the study analysis(Edwards et al. 2018). Our data from the pilot suggests that the prevalence of missing data is similar to that in other studies and we would expect a similar level of missing OKS and OHS data in the main POSt-Plus study. This demonstrates that missing OKS and OHS data is a significant issue and this chapter has explored the different ways of mitigating against this.

The best way to manage missing data is to ensure that the study is designed and conducted in a manner that helps minimise missing data. Table 14 describes the processes that have been put into place for POSt Plus, as a result of the POSt Pilot, to help reduce the levels of missing data in the main study.

Table 14: Recommendations for minimising missing data utilised in POSt Plus as a result of the POSt-Pilot study(Mercieca-Bebber et al. 2016)

Design	Subject	Recommendation	Disadvantages
Assessment schedule	Specific assessment time points	Specify minimum data requirements for PROMS required e.g baseline, 6 months	Can create impression that additional PRO assessments are unimportant
	Align PROM assessment with clinic visits	Align PROMs assessments to clinic/hospital visits so that data may be captured while the patient is in hospital	Burdensome to participants to attend regular assessments
	Shorter follow up duration (patients more likely to drop out of a study the longer the follow up study time period)	Timepoints used for follow up: preoperative (baseline); immediate postoperative (24-48h) whilst participant is still an inpatient in hospital and forms can be collected in hospital; 3 months and 6 months.	May lose important data if the follow up time period is too short
	Define PROM assessment time windows	Specify the time periods exactly in the study protocol (e.g 6-months post-operatively: This is defined as the time period of between 5 months and 8 months after surgery)	The time windows may not be flexible
Additional data collection	Auxiliary data to assist in data analysis and interpretation	Collect additional data (other than primary PROMs) such as clinical data, patient demographics, comorbidities, health quality assessments.	Additional resources and time to collect data and risk of overburdening patients
	Record reasons for missing data (See Figure 7)		
Eligibility criteria	Inclusion and exclusion criteria	Inclusion criteria: participants must be able to complete questionnaires independently Exclusion criteria: exclude participants who cannot understand or read the questionnaire in English or have cognitive impairment	Ability to complete PROMs may change during course of treatment and study time period, especially in the elderly or over a long follow up period May not be generalisable to all patients May reduce sample size and power of study

	Baseline PROM data completeness as an eligibility criterion	Only include participants who have a complete baseline PROMs dataset	Risk of selection bias May not be generalisable to all patients
Feasibility of study	Pilot study	Determine feasibility, resources, compliance, acceptability and potential issues by conducting a pilot study Use pilot study as a training opportunity for research staff	
	PROM resources	Dedicated research member responsible for distributing and collecting PROMs data and checking for data completeness and accuracy and following up participants.	Requires resources and time and funding. Can be difficult to recruit research staff
PROM training and guidance for research staff	PROM guidance and support	Provide detailed SOP for research team members to ensure consistency in study conduct. Provide training for research team to ensure strict adherence to timepoints and SOPs of study protocol.	
MOA for PROMs (eg hard copy, electronic, telephone)	Postal MOA With flexible MOA	Complete baseline assessment in clinic and subsequent assessments by post. Use postage-paid self-addressed envelope to facilitate easy return of completed questionnaires. Use alternative MOA if participant fails to complete PROMs fully, such as a telephone call to fill in missing PROMs data	Requires additional staff time and postage costs May be burdensome for participants to send questionnaires back to researchers Questionnaires can be lost in postal system Telephone answers may be biased as participants try and please researcher and requires a flexible assessment time window
Minimising burden of PROMs	Offer assistance to participants to complete PROs (to reduce burden PRO completion)	Try to avoid additional clinic visits where possible	
	Questionnaire content	Provide clear and simple instructions to complete PROMs Ensure format of PROMs is clear and uses large font Avoid multiple repetition of questionnaires Use consistent uniform presentation of PROMs	Environmental burden Increased printing costs due to additional pages in the questionnaire booklet May not be possible if using more than one questionnaire

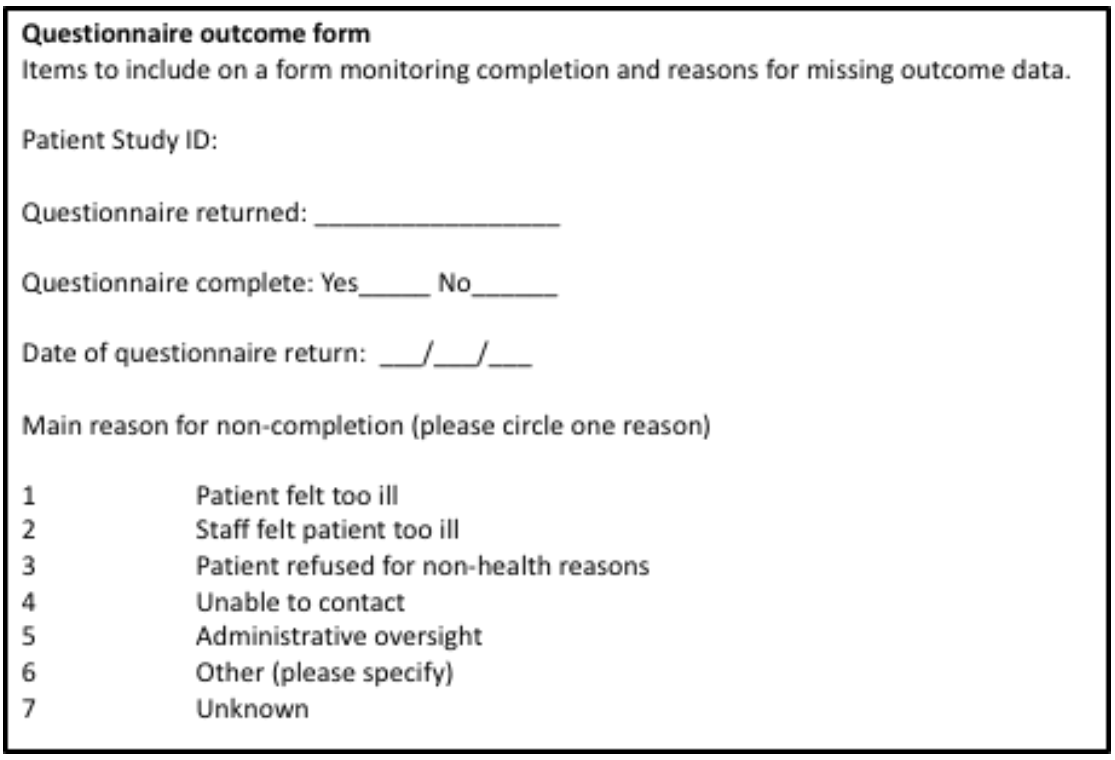
		Consider single sided PROMs (as some participants may forget to turn over and complete the other side of the questionnaire(s)	
	Use validated questionnaires	The validation process for the PROMs will ideally have addressed some of the issues associated with questionnaires, making completing the PROMs easier for participants	May incur costs for using certain validated questionnaires and requires permission(s) Must ensure that validated questionnaires used are formatted exactly as they were intended to be used
Include PROMs in SAP	Specify potential problems with PROMs analysis in SAP	Statistical analysis plan for dealing with missing data	May not be possible to predict and mitigate for all potential problems in SAP. The pilot study can help with this.
	Plans for addressing missing data in SAP	Statistical analysis plan for dealing with missing data	May not be possible to fully plan how missing data will be handled prospectively. The pilot study can help with this.
Sample size	Increased sample size to allow for participant attrition	Build in a larger sample size than is required to allow for missing data	Whilst increasing the sample size can improve the study power, it cannot mitigate entirely against the level of missing data in participants. The outcomes of participants with missing PROMs data may be different from those who have complete data, leading to bias.
Team to design study protocol	Involve MDT to assess study protocol, design and conduct	Involve experienced investigators, research nurses and patients to review study protocol to identify areas for maximising patient compliance and adherence to study processes.	None
Administration process	Approach all participants	All participants involved in the PROMs study should be approached to complete scheduled PROMs assessments to avoid selection bias	Requires dedicated PROMs research team member and extra resources and time.
	Organised	Ensure sufficient questionnaires available for use. Ensure questionnaires sent out in time and contact participants by phone to remind them that the questionnaires are about to be sent out and to check that they have received them. Prepare to handle potential problems.	Requires dedicated PROMs research team member and extra resources and time.

		Keep track of when PROMs are due at the different timepoints.	
	Checking	Checking source data (data entry of questionnaire data into database). Digitally scan PROMs data so original questionnaire is saved and can be used to verify data entry accuracy and completeness. Check questionnaires as they are returned, as soon as possible for completeness and errors, so participants can be contacted to complete any missing data or clarify reasons for missing data and record these reasons.	Requires dedicated PROMs research team member and extra resources and time.
	PROM completion cover sheet	Use cover sheet which records for each participant, whether or not PROMs have been completed in full and if not, the reasons why. Standardised reasons for missing data.	Reasons for missing PRO data may not be easy to determine in some cases. Requires dedicated PROMs research team member and extra resources and time.

PROMs: Patient reported outcome measures; SOP: Standard operating procedures; MOA: Mode of administration; SAP: Statistical analysis plan

All of the above recommendations were put into place for the main POSt Plus study to try and minimise the amount of missing data present in the larger study and closely follows the recommendations from several authors on study design and methods of reducing missing data in clinical research(Mercieca-Bebber et al. 2016; National Research Council et al. 2011; Ibrahim et al. 2012; Wisniewski et al. 2006; Little, D'agostino & Cohen 2012b). Despite this, missing data will still occur and the next step in how to manage this is to explore ways of using different statistical models to analyse the data.

Figure 7: Example of an outcome form for questionnaires, recording the reason for non-completion of a questionnaire(Bell & Fairclough 2014)



Questionnaire outcome form

Items to include on a form monitoring completion and reasons for missing outcome data.

Patient Study ID:

Questionnaire returned: _____

Questionnaire complete: Yes_____ No_____

Date of questionnaire return: ___/___/___

Main reason for non-completion (please circle one reason)

1	Patient felt too ill
2	Staff felt patient too ill
3	Patient refused for non-health reasons
4	Unable to contact
5	Administrative oversight
6	Other (please specify)
7	Unknown

Determining the type of missing data can be challenging and influences the type of statistical analysis that can be undertaken to deal with the missing data. Unfortunately, one of the biggest difficulties lies in being unable to differentiate for certain whether missing data is MNAR or MAR as this is based solely on the data that is available(Ibrahim et al. 2012). As a result, there is a growing consensus that a

sensitivity analysis should be undertaken on the available data to try out different missing data mechanisms to examine how sensitive the results are to the assumptions of whether missingness is MNAR(Ibrahim et al. 2012). Although as part of the pilot study and this thesis, a sensitivity analysis was not undertaken, one of the recommendations for the main study going forward would be to include a sensitivity analysis as part of the analysis plan for dealing with missing data in POSt-Plus.

There are two main methods of undertaking a sensitivity analysis in order to provide a MNAR modelling framework(Leurent et al. 2018). The selection model method specifies the conditions in which the data is observed (or “selected”) as a result of the underlying data values(Leurent et al. 2018). One assumes that the odds of a non-response change is known and changes with the values of an outcome. The probability of a nonresponse (missing data occurring) is a function of the outcome and there is a specific (unverifiable) constant linking the two(National Research Council et al. 2011). For instance, the chance of being missing doubles for every decrease of 1% in quality of life. The advantage of the selection model is it can be included as part of the analysis model for missing data such as using an inverse probability weighting approach(National Research Council et al. 2011). An example of this approach described by Carpenter et al. used an MI model under the MAR assumption in order to obtain parameter estimates for each imputed data set followed by a selection model sensitivity analysis to produce an overall MNAR parameter estimate, which is a weighted average of these parameter estimates, where the weights depend on an assumed degree of departure from MAR(Carpenter et al. 2007). The biggest drawback of the selection model is that one cannot verify the assumptions about the conditional distribution of the unobserved data and it is sensitive to any departure from the assumptions(Leurent et al. 2018).

The other approach involves a pattern mixture model in which the MNAR issue is formulated in terms of the different distributions between the missing and observed data. For example, the participants with missing data have a 1% worse quality of life outcome compared to those observed with complete data. The overall distribution of a variable is seen as a mixture of the distribution of the observed and the distribution

of the missing values ('pattern-mixture') (Leurent et al. 2018). A big advantage of this model is that it can be easily incorporated into a missing data analysis model such as MI to reflect different possible deviations from an MAR assumption. It involves the following steps(Leurent et al. 2018):

6.4.1 Use MI to impute the missing values under an MAR assumption

- Modify the MAR-imputed data to reflect a range of plausible MNAR scenarios, for example, by multiplying the imputed values by a specific constant *c* or by adding a constant δ.
- Analyse the resulting dataset as planned with the multiply-imputed dataset, fitting the analysis model to each imputed dataset and combining the results using Rubin's rules.

The Oxford group have published extensively on rules governing the scoring of the OKS and OHS. In particular, when it comes to missing data, the authors have suggested that if after repeated attempts to obtain complete data from an individual, only one or two questions have been left unanswered, it is reasonable to enter the mean value representing all of their other responses, to fill the gaps(Murray, Fitzpatrick, Rogers, Pandit, Beard, Carr & Dawson 2007b). This does not appear to be based on any specific statistical analysis but rather on a "rule of thumb" based on common sense and practical clinical application in a real-world setting. Our simulation work in this Chapter sought to compare how using these rules fared against a MI model when one or two items were missing from the OKS. Table 6 shows that in such a situation, with one item missing, there was not much difference between the Oxford rules and MI model although the mean difference using MI was smaller. When 2 items were missing, the mean difference in total OKS was less with the MI model compared to the Oxford rules model but again, unlikely to be clinically significant as the difference was approximately by 0.5 point. Although the Oxford rules state that if 3 items or more were missing, then the total score could not be calculated, when we tested the Oxford rules in this scenario with the MI model, the mean difference in scores was again approximately 0.5 points between the 2 models, although the MI model was more precise. This is to be expected given that the MI model, even in this small sample sized simulation study, uses other predictor variable to impute the missing data. One

would expect that if the MI model sample size was greater, with more predictor variables incorporated into the model, then the accuracy of the modelling would be greatly enhanced. The conclusion from this simulation exercise is that the Oxford rules appear to be a good practical “rule of thumb” way of calculating the total OKS score even if one or two items were missing and that MI modelling in a clinical setting is probably not warranted. It could also be argued that even with three items missing, the Oxford rules may still be justified although MI modelling may provide a greater degree of accuracy. It remains to be seen whether once four or more items are missing whether the Oxford rules are valid and, in this situation, MI modelling may still allow the valuable remaining data that is available to be used without discarding it.

6.4.2 Limitations

There are limitations with this simulation pilot study that must be acknowledged. Firstly, the sample size of the simulation exercise was small. There were only 37 patients in the knee group dataset used in the simulation models. This meant that the number of predictor variables that could be incorporated into the MI model had to be restricted to just three. Any more led to issues of non-convergence in the regression models. It would appear reasonable to suggest that with a larger sample size (n> 200), that all the items of the OKS where data was available could be built into the MI model and predictor variables, which would enable the imputation model to be more accurate(Rombach et al. 2018).

A second potential limitation is that an assumption has been made that the missing data is MAR and not MNAR. For the reasons given previously, we would recommend undertaking not only an imputation analysis of the missing data but also suggest incorporating a sensitivity analysis (pattern mixture model) in order to account for the possibility that the missing data is MNAR. Having said this, it is likely that the pattern of missing data shown in our simulation study whereby whole sections of items are missing are likely to be MCAR mainly as a result of the participant not receiving the questionnaires or forgetting to turn over the forms to complete the other side of items. Participants who have only a few items missing are assumed to have MAR data.

Non-convergence (perfect prediction) of the imputation models was a third limitation. With the MI model, the issue of non-convergence occurred as the predictor variables were also found to have missing data in the simulation model in this small sample size. With a larger sample size, other simulation models with 4 items, then 5 items then 6 items missing could be tested to see the effect of MICE on imputing this missing data. This problem has been recognised especially with smaller sample sizes and greater levels of missing data, with imputation modelling being performed at the item level. In a large study looking at the imputation modelling for missing data in the Oxford knee scores, Rombach et al. found that for small sample sizes combined with a large proportion of missing data at the item level, then non convergence was likely to be a significant issue(Rombach et al. 2018).

The difference between imputing at the composite score (unit) level or at the item/subscale level has been shown to be small. Imputing at the item level is likely to provide more precise estimates of treatment effect as it takes into account the correlation between the different items compared to imputing at the composite score level(Rombach et al. 2018). It is however more prone to problems with non-convergence as shown in the pilot study here.

6.4.3 Recommendations

The pilot study has identified a number of ways of improving the design and conduct of the main study to help limit the amount of missing data that occurs (Table 10). The results of the pilot study have shown the pattern of missing data likely to be observed and demonstrated that an MI approach is feasible for mitigating against missing data in the main study, when more than 3 items are missing but using the Oxford rules is reasonable if only one or two items are missing (See Figure 3). The model chosen here was imputing at the item level but composite score (unit level) imputing should also be performed with the larger sample size available in the main study to see if there is any difference in estimations of treatment effect. If issues of non-convergence continue to occur despite the larger sample size, then unit level imputing may be more feasible rather than item level. Furthermore, further work is needed to see how many items could be used in the MI model as there may be a limit. A sensitivity analysis

should also be performed in order to test the assumption that the missingness in the data is MAR and not MNAR. Finally, using multiple imputation by chained equations (MICE) for the larger POSt Main study is another method that should be considered when analysing missing data in the main study dataset. The pilot used a complete dataset in order to test the role of MI in handling missing data when a single variable had some missing data incorporated into the model. In reality, the likelihood is that the main study will have several variables containing some missing data and this is where MICE can be a useful tool for handling such situations.

Missing data can occur in several variables especially in larger datasets and the use of MICE is a useful and practical method of generating multiple imputation models, one for each variable that has missing data present(Royston & White 2011b; White et al. 2010). The alternative nomenclature for MICE is fully conditional specification or sequential regression multivariate imputation.

In essence, MICE is performed with the following steps(Royston & White 2011):

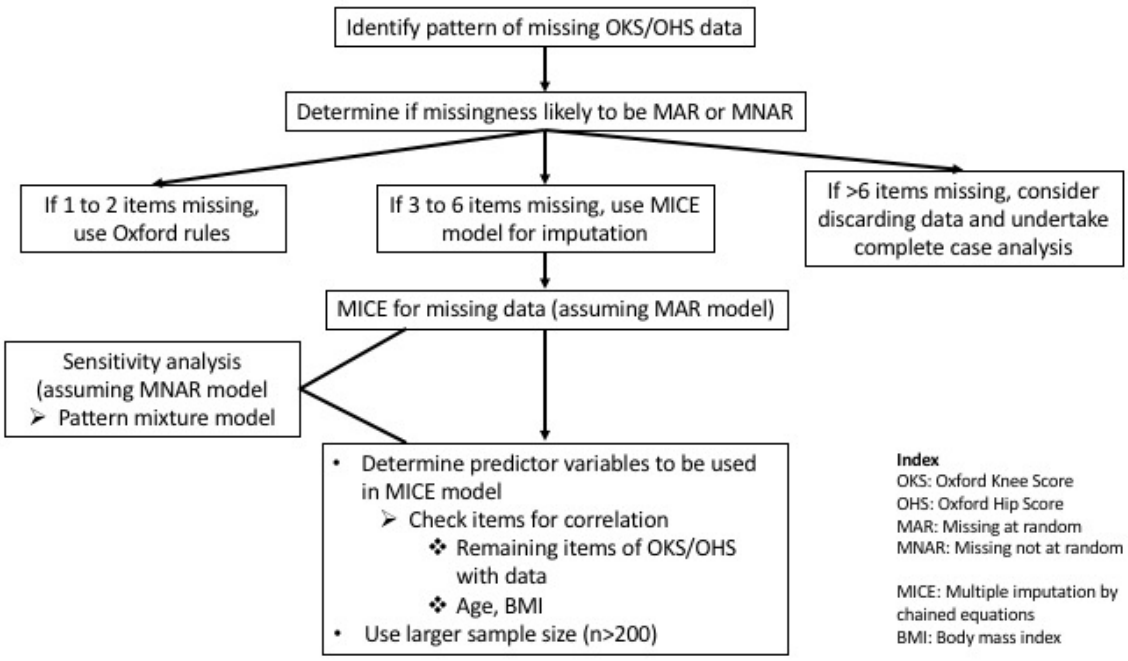
- All missing values are filled with randomly generated values by sampling the observed values.
- The first variable with missing values is then regressed on all other variables. Each of the variables may contain missing observations. The estimation is restricted to participants who have observed values for Variable 1 whilst the missing values in Variable 1 are replaced with simulated plausible values drawn at random from the observed values of Variable 1 (using posterior positive prediction distribution).
- The next variable (Variable 2) has missing data and is regressed to all other variables. Step 2 is repeated again and again for Variables 3, 4, 5 and so on, for a number of cycles (*n*) in order to produce a single imputed dataset.
- By convention, at least 10 cycles are required in order to produce convergence of the sampling distribution of imputed values(Royston & White 2011; van Buuren 2007). The imputed values at the end of the 10th iteration, combined with the observed data, will make up one imputed data set. The whole process is repeatedly *M* times to produce *M* number of imputed datasets.

The assumptions for MICE are identical to that required for any MI method. There is an assumption the data is missing at random (MAR), that as many predictor variables should be fed into the imputation model so as to minimise bias and the analysis model must be appropriate, for example using linear regression for continuous data. Rubin's rules are required in order to ensure the correct approach is used, combining estimates of interest (e.g., regression coefficients) across the *M* imputed datasets.

There are several advantages in using MICE for imputation modelling.

- It is a very flexible technique, allowing each variable to be modelled individually.
- The data can be different (binary, categorical, continuous) and the distribution can be varied (Gaussian, Poisson etc.).
- It can handle monotone missing data patterns as well as non-monotone and arbitrary missing patterns.
- Subset analysis can be performed by setting boundaries such as only imputing data in female patients or over a certain age for example.
- It can include in the imputation model variables that are functions of other variables

Figure 3: Recommendation for statistical analysis plan for handling missing OKS and OHS data



6.5 Conclusion

The best way of minimising the occurrence of missing data in the main study is to follow the recommendations described above, in terms of study design and conduct. If despite this, missing data occurs (which is inevitable) then at least the amount of missing data present would be smaller and easier to handle using multiple imputation to provide accurate estimates of treatment effects. As part of this process, a sensitivity analysis should also be conducted in order to challenge the underlying assumptions about the type of missing data mechanisms. When undertaking MI in the main study, MICE should be considered as the process.

Chapter 7 The use of accelerometers

7.1 Introduction

Physical activity (PA) is the body movement produced as a result of the expenditure of energy by skeletal muscle(Caspersen et al. 1985). There is strong evidence that improving PA in the population has overall health quality benefits in terms of managing obesity, hypertension and other conditions(Mok et al. 2019). Sedentary behaviour is thought to be one of the reasons for the rising obesity epidemic in many developed nations(Mok et al. 2019). Regular PA is also recommended for those with osteoarthritis but only a small proportion of patients meet the recommended levels of PA, defined as 7000-10,000 steps/day(Department of Health 2011; Wallis et al. 2013). The impact of pre-optimising PA in patients awaiting THA or TKA remains unclear. A systematic review demonstrated benefits of exercise programmes for those waiting THA but no benefit for those waiting for TKA(Gill & McBurney 2013); the authors suggested the majority of exercise programmes implemented were below the minimum intensity threshold to sufficiently challenge the body(Gill & McBurney 2013). These studies investigated structured programmes of exercise and little is known on the relationship between daily activity patterns, including sedentary behaviour and non-intentional activity. The preoperative level of physical activity has been reported to influence outcomes related to self-reported disability, daily physical function and quality of life(Hendrick et al. 2009). Studies have depended on patient self-reports of physical activity, which may be subject to bias, particularly in those people with low mood and those with lower levels of psychological wellbeing(Hendrick et al. 2009). A more objective assessment of physical activity is required, and this can be conducted using accelerometry(Mathie et al. 2004; Yang & Hsu 2010).

There are two ways of measuring PA; subjective and objective methods. Subjective methods include the use of physical activity questionnaires such as the International Physical Activity Questionnaire (IPAQ), activity diaries and survey. These have the advantage of being low cost and easy to utilise. They have the disadvantages of being

subjective, as they are self-reported, and therefore can lead to inconsistent results(Yang & Hsu 2010). Over the last few years, attention has turned to the use of wearable or body-fixed motion capture devices, including pedometers, accelerometers and even mobile phones and watches with built in activity monitors. Pedometers are the simplest devices that can be used to collect PA data, in the form of the number of steps taken, using a spring-loaded mass or switch system, that registers every time a step is taken. Whilst cheap and easy to use, pedometers only provide basic data on PA and does not provide any information on the intensity of PA(Yang & Hsu 2010). Accelerometers are devices that measure the acceleration of an object in motion along lines of axes(Yang & Hsu 2010). As acceleration is proportional to an external force, data from accelerometers can reflect the intensity and frequency of human movement and can be used to derive velocity and displacement information with respect to time(Chen & Bassett 2005). Most modern accelerometers contain piezoelectric sensors that measure acceleration in one to three orthogonal planes (anteroposterior, mediolateral, and vertical) (John & Freedson 2012; Chen & Bassett 2005). The processed data is then recorded and saved by the internal memory of the accelerometer and then downloaded to a computer either through a USB cable or wirelessly by Bluetooth technology.

The importance of performing physical exercises following arthroplasty is well known(Pozzi et al. 2013), but there are limited studies on the effect of daily physical activity on outcome after joint replacement surgery(Paxton 2015; Walker et al. 2002). One relatively small study demonstrated a correlation between increased post-operative physical activity measured by accelerometry, and improved physical activity and less disability at six months follow-up(Lareau 2008). One of the challenges lies in the fact that different studies use different methods of measuring PA, both subjective (questionnaires) as well as wearable devices (eg. accelerometers) which leads to inconsistencies in the literature. This is further compounded by differences in the patient population being assessed and the duration of follow up(Paxton 2015). Table 1 shows a summary of the most recent studies in which accelerometers were used with or without accompanying outcome questionnaires, in order to assess PA after TKA and/or THA. The studies were mainly longitudinal and prospective in nature and all were relatively small in size involving less than 100 subjects. A range of different

accelerometers and outcome questionnaires were used illustrating the difficulty in achieving consistency in this topic.

Accelerometers provide a wealth of data which is recorded continuously but it is important to identify one or two outputs which best represents PA. Some studies have focused on the number of steps per unit time as a marker of PA whilst others have used the intensity of activity (e.g sedentary/light/MVPA activity) as the outcome measure(Skender et al. 2016). The metabolic equivalent of task (METs) is an objective measure of the ratio of the rate at which a person expends energy, relative to that person's body mass, while performing a specific physical activity, compared to a reference (set by convention at 3.5 ml of oxygen per kilogram per minute). This has also been used as a measure of PA(Skender et al. 2016). Clearly, the type of output used in a study will depend on the type of accelerometer, with the more sophisticated and modern accelerometers able to collect more data.

Patients often report that their mobility and function is significantly improved after THA and TKA. Several studies, as shown in Table 1, suggest that patients believe they are more physically active after their surgery, based on self-reported questionnaires(Paxton 2015). However, this is not reflected in the objective measurement in PA(Franklin et al. 2006; de Groot et al. 2008; Hayes et al. 2011; Vissers et al. 2013; Harding et al. 2013; Thewlis et al. 2019). Other studies contradict this perspective and patients show an improvement in both self-reported outcome questionnaires and accelerometer data(Walker et al. 2002; Tsonga et al. 2011; Jiang et al. 2015; Höll et al. 2018).In the only longer term study of its kind, at up to four years, Vissers et al. showed that PA did not improve after TKA or THA as measured by acclerometers and indeed, the daily activity level actually reduced at four years compared to at six months postoperatively. This is despite the fact that patient perceived that their physical function had improved, based on self-reported questionnaires(Vissers et al. 2013).

Table 1 Summary of studies that used accelerometers +/- questionnaires to assess PA in TKA and THA patients

Reference	**Study type**	**Surgery**	**Publication year**	**No. subjects**	**Type of assessment**	**Assessment**	**Follow up Duration****	**Change in PA/function**
Walker et al.(Walker et al. 2002)	Longitudinal, prospective	TKA	2002	19	SRQ; Acc	Acc (Numact) NHP	6m	SRQ: Improves then deteriorates after 3m Acc: Increases
Franklin et al.(Franklin et al. 2006)	Longitudinal, prospective	TKA	2006	31 (8*)	SRQ; Acc	Acc (SAM) Activity log SF-12	6m (6wk*)	SRQ: Improves Acc: Decreases
De Groot et al.(de Groot et al. 2008)	Longitudinal, prospective	TKA;THA	2008	80 TKA 44 THA 36	SRQ; Acc	Acc (IDEEA) PASIPD	6m	SRQ: Improves Acc: No change
Vissers et al.(Vissers et al. 2010)	Longitudinal, prospective	TKA	2010	44	SRQ; Acc	Acc (AM monitor) SF-36	6m	SRQ: Improves Acc: No comparison as no preop data
Hayes et al.(Hayes et al. 2011)	Cross sectional	TKA	2011	65	SRQ; Acc	Acc (IDEEA) OKS; Tegner	12m	SRQ: Improves Acc: No change

Tsonga et al.(Tsonga et al. 2011)	Longitudinal, prospective	TKA	2011	52	SRQ; Acc	Acc (Digiwalker Yamax) SF-36; PASE	6m	SRQ: Improves Acc: Increases
Brandes et al.(Brandes et al. 2010)	Longitudinal, prospective	TKA	2011	53	Acc	Acc (SAM, DynaPort ADL monitor)	12m	Acc: Increases
Krenk et al.(Krenk et al. 2013)	Longitudinal, prospective	TKA;THA	2013	20 TKA 8 THA 12	Acc	Acc (Actiwatch Spectrum)	Up to 9d	Acc: Decreases
Vissers et al.(Vissers et al. 2013)	Longitudinal, prospective	TKA;THA	2013	44 TKA 23 THA 21	SRQ; Acc	Acc (AM monitor) HOOS/KOOS	4y	SRQ: Improves Acc: Decreases
Harding et al.(Harding et al. 2013)	Longitudinal, prospective	TKA;THA	2014	57 TKA 33 THA 24	SRQ; Acc	Acc (ActiGraph GT1M); OHS; OKS; UCLA	6m	SRQ: Improves Acc: No change
Jiang et al.(Jiang et al. 2015)	Longitudinal, prospective	TKA	2015	50	SRQ; Acc	Acc (Fitbit) WOMAC, KSS	6m	SRQ: Improves Acc: Increases

Reference	Study type	Surgery	Publication year	No. subjects	Type of assessment	Assessment	Follow up Duration**	Change in PA/function
Toogood et al.(Toogood et al. 2016)	Longitudinal, prospective	THA	2016	33	Acc	Acc (Fitbit)	4 wk	Acc: Increases
Höll et al.(Höll et al. 2018)	Longitudinal, prospective	THA	2018	46	SRQ; Acc	Acc (SAM) WOMAC; HHS	3m	SRQ: Improves Acc: Increases
Thewlis et al.(Thewlis et al.2019)	Longitudinal, prospective	THA	2019	51	SRQ; Acc	Acc (GeneActiv) HOOS	6m	SRQ: Improves Acc: No change

* accelerometry data **d: days; wk: weeks; m: months; y: years PA: Physical activity; SRQ: Self-reported questionnaires; Acc: Accelerometer;

Questionnaires: PASE: Physical Activity Scale for the Elderly; SF-12: Short Form Health Survey-12; SF-36: Short Form Health Survey-12; UCLA activity: University of California Los Angeles Physical Activity Questionnaire; PASIPD: Physical activity scale for individuals with physical disabilities; OHS: Oxford Hip Score; OKS: Oxford Knee Score; HOOS: Hip Disability and Osteoarthritis Outcome Score; KOOS: Knee Injury and Osteoarthritis Outcome Score; NHP: Nottingham Health Profile questionnaire; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index; HHS: Harris Hip Score; KSS: Knee Society Score;

Activity Monitors/Accelerometers: IDEEA: Intelligent Device for Energy Expenditure and Activity (Minisun, Fresno, USA); SAM: Step-Watch 3TM Activity Monitor (OrthoCare Innovations); Actigraph (ActiGraph LLC, Fort Walton Beach, FL, USA); Actiwatch Spectrum (Philips Respironics, 1010 Murry Ridge Lane, Murrysville, PA 15668, USA); AM monitor (Temec Instruments, Kerkrade, The Netherlands); Numact monitor (Newcastle); GeneActiv (Cambridge, United Kingdom); Fitbit (Fitbit Inc; USA)

The increasing use of wearable devices to measure PA means that further research is needed into their efficacy and whether or not they can help improve patient outcomes after THA and TKA. In a review article, Bahadori et al. suggested that such research remains limited and the five studies included in this review were at risk of significant bias, leading the authors to conclude that at present, the widespread routine use of accelerometers and other wearable devices in clinical care was not supported(Bahadori et al. 2018).

The overall objective of this Chapter was to investigate the use of accelerometers in the POSt Plus study. The primary aims of this pilot study were to:

- identify one or two specific outputs from accelerometer data that would be used as markers of PA in any subsequent analysis.
- assess if PA changes after TKA or THA using an accelerometer and if this data was also captured using the OKS/OHS.
- identify if there was any correlation between objective measurements of PA using accelerometers and the OHS/OKS.
- identify if there were any practical issues using the accelerometers, in terms of loss of the accelerometers, complications related to wear and ease of use by patients.
- provide a recommendation on the use of accelerometers in the main study POSt-Plus.

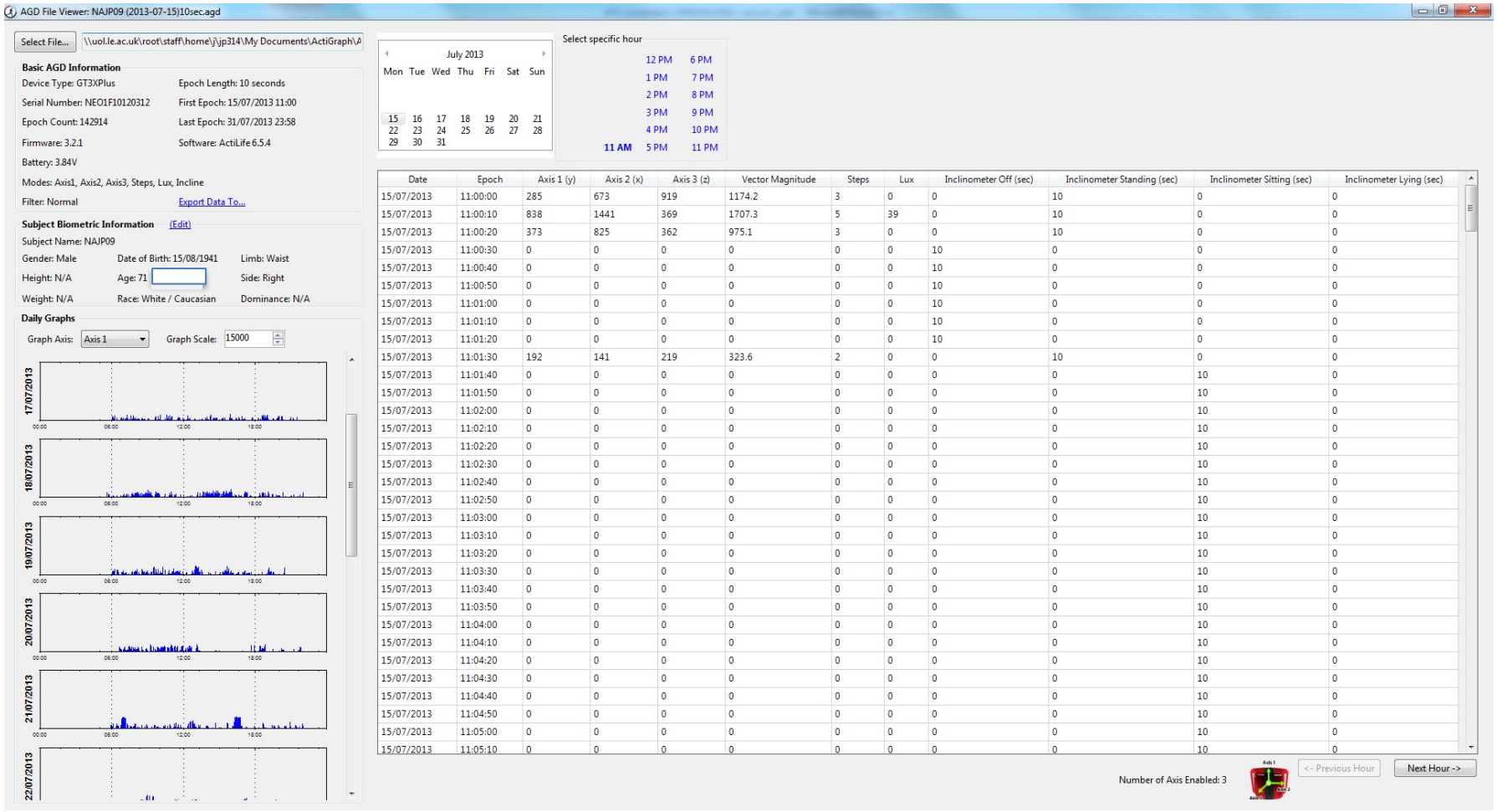
7.2 Methods

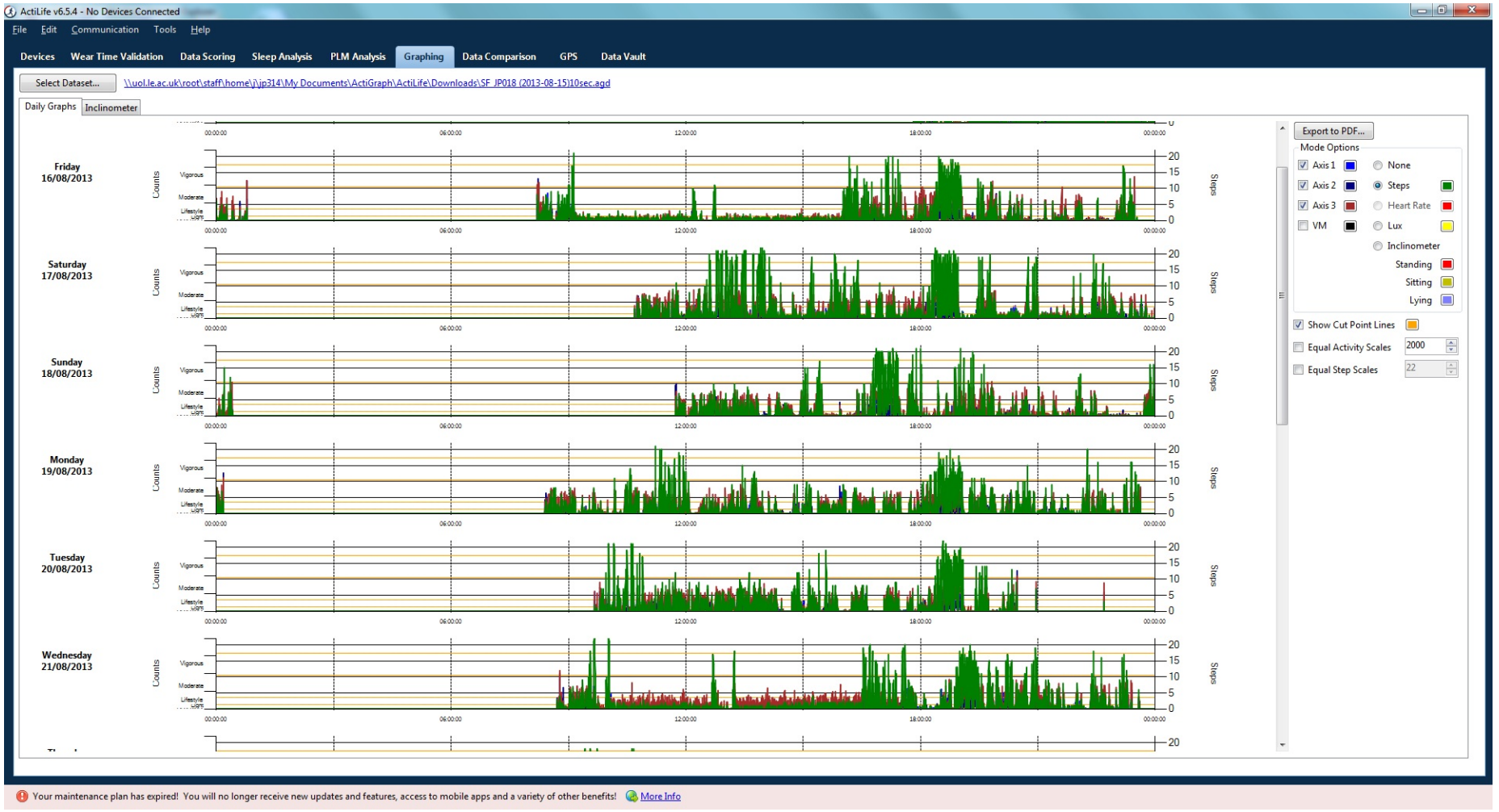
The primary outcome of interest was physical activity (PA), measured using the ActiGraph GT3X (AG3X) activity monitor (ActiGraph LLC, Fort Walton Beach, FL, USA). It contains a triaxial accelerometer and is a small, lightweight device worn around the waist under or over clothes (see Figure 1). The AG3X accelerometers are validated for use in clinical studies to assess PA and especially sedentary activity in subjects[13-16]. Accelerometers were initialized as

per the manufacturer's manual. The epoch (the summation of the frequency and intensity of accelerations and decelerations measured during the select time intervals) was set at 10 seconds. Participants were instructed to attach the accelerometer as close as possible to the body's centre of mass, over their right hip, using an elastic belt around their waist and to wear it for a minimum of 10 hours, during waking hours, for 7 consecutive days (minimum of 3 days). All participants also received detailed instructions on how to wear the accelerometers and how to look after them using an accelerometer information sheet (Appendix 4). Data from activity monitors was downloaded to a computer with the ActiGraph software (ActiLife) installed. The minimum requirement for nonzero epochs within an hour was set at 5%. Therefore, for hourly activity data to be valid, a minimum of 3 minutes of nonzero epochs per hour had to be recorded. Furthermore, for a data set to be valid, the activity monitor had to be worn for at least 10 hours for a minimum of 3 of the 7 days. Wear time calculations using ActiLife were performed only on valid data sets. Patients were asked to wear the accelerometer for 7 days prior to admission for their operation and 7 days at their 3-month postoperative time period. Patients were also asked to complete a number of questionnaires including the OHS and OKS pre-operatively and at three months post-operatively. Examples of the outputs from the Actigraph GT3X are shown in Figure 2.

Figure 1: (A) Actigraph GT3X (B) Actigraph GT3X worn over waist right hip

Figure 2: (a) Example of the activity output from Actigraph GT3X and (b) daily wear time of participants




ActiLife v6.5.4 - No Devices Connected
File Edit Communication Tools Help
Devices Wear Time Validation Data Scoring Sleep Analysis PLM Analysis Graphing Data Comparison GPS Data Vault
Select Dataset...
\\uol.le.ac.uk\root\staff\home\j\jp314\My Documents\ActiGraph\ActiLife\Downloads\SF_JP018 (2013-08-15)10sec.agd
Daily Graphs Inclinometer
Friday 16/08/2013
Saturday 17/08/2013
Sunday 18/08/2013
Monday 19/08/2013
Tuesday 20/08/2013
Wednesday 21/08/2013
Counts
Steps
Vigorous
Moderate
Lifestyle
Light
00:00:00
06:00:00
12:00:00
18:00:00
Export to PDF...
Mode Options
Axis 1
Axis 2
Axis 3
VM
None
Steps
Heart Rate
Lux
Inclinometer
Standing
Sitting
Lying
Show Cut Point Lines
Equal Activity Scales
2000
Equal Step Scales
22
Your maintenance plan has expired! You will no longer receive new updates and features, access to mobile apps and a variety of other benefits!
More Info

Data was analysed using Stata Version 13.1 (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP). Data was checked for normality and was found to be not normally distributed thus the non-parametric Wilcoxon signed rank test was used in Table 1 to test for significance. Probability values (p) of < 0.05 were considered statistically significant. The Spearman test was used to test for correlation. The strength of the Spearman correlation coefficient was determined using the following guide for the absolute value of rho (ρ): 0.00-0.19 “very weak”; 0.20-.39 “weak”; 0.40-0.59 “moderate”; 0.60-0.79 “strong”; 0.80-1.0 “very strong”.

7.3 Results

Complete datasets of validated pre- and post-operative accelerometer data was available in 73% of participants (66 patients out of the cohort of 90 patients) and patient demographic data and OHS/OKS available in 76 patients (84%). The patient demographics are shown in Table 2. The patients who had THRs were slightly younger compared to patients who had TKRs on average and there were more women in this group compared to men. The mean BMI in the knee patient group was also higher than that in the hip patient group.

Table 3: Patient demographics for the knee and hip replacement patient groups. Where appropriate, the mean data is presented with standard deviation (SD).

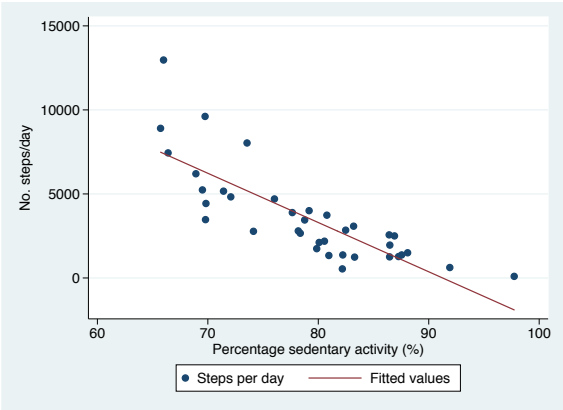
Demographic	**Knee (n=37)**	**Hip (n=39)**
Age (years)	70.4 (9.1)	67.1 (7.5)
Gender		
Male	19	11
Female	18	28
ASA		
1	1	4
2	30	31
3	6	4
BMI	32.8 (7.7)	29.8 (6.2)

7.3.1 Which accelerometer output(s) to use

Two outputs from the accelerometers were chosen as surrogate markers of PA, based on a review of the literature (see Table 1). These were the number of steps recorded daily and the percentage time spent in sedentary activity. Both these outcome measures showed a very strong correlation with one another. Although the scatter plots show a linear correlation between two variables, in reality, this would not be true and there would be a curvilinear correlation instead. The line presented in the graphs below represent a best fit line rather than the true curvilinear correlation line but for the purposes of demonstrating if there was a correlation (positive or negative), the best fit line would suffice.

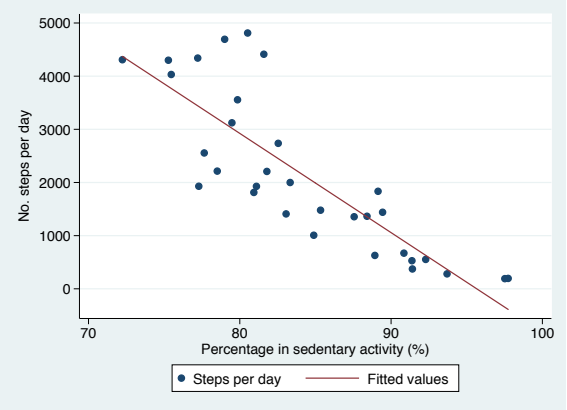
Figure 3: Scatter plot of number of steps/day and percentage time spent in sedentary activity in TKA patients

(a) Preoperative



Spearman correlation: -0.8721

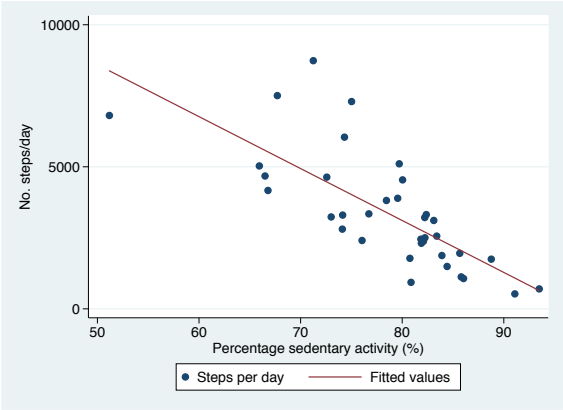
(b) Three month post-operative



Spearman correlation: -0.8644

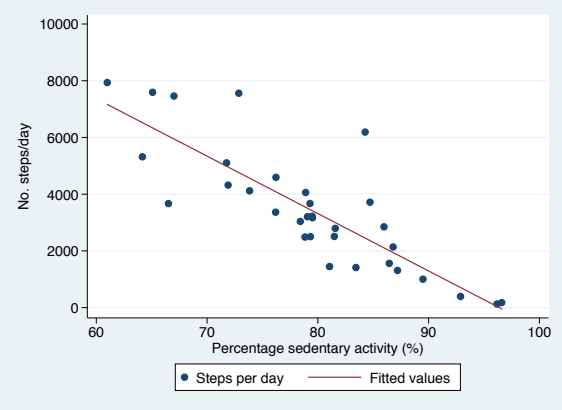
Figure 4: Scatter plot of number of steps/day and percentage time spent in sedentary activity in THA patients

(a) Preoperative



Spearman correlation: -0.8118

(b) Three month post-operative



Spearman correlation: -0.8212

7.3.2 Does PA change after TKR or THR?

There was a significant decrease in physical activity, as measured by the number of steps taken per day and percentage time spent in sedentary activity, in knee replacement patients. Patients spent almost all of their time (97.5%) in either sedentary or light activity but only spent 2.1%of their time in moderate to vigorous physical activity (MVPA). After their knee replacement, at three months, there was a significant decrease in physical activity as shown in Table 1. The mean number of daily steps taken decreased significantly from 3191 to 2133 (p=0.0047) and percentage time spent in sedentary activity increased significantly from 79.4% to 85.1% (p=0.0008).

For hip replacement patients, the mean number of daily steps taken increased from 3460 to 3599 (p=0.6946) and percentage time spent in sedentary activity remained unchanged 78% (p=0.6808). Hip replacement patients also spent 98.1% of their time in either sedentary or light activity. Only 2% of their time was spent in MVPA. Mean OKS/OHS did improve significantly however from

16 to 35 (p=<0.0001) and from 16 to 37 (p=<0.0001) respectively. Table 3 summarises the findings of the amount of physical activity and OKS/OHS results at the preoperative and three month postoperative time period.

Table 4: Amount of physical activity (intensity and number of steps/day) and OKS/OHS preoperatively and at three months postoperatively

Knee					
Physical activity	n	Preoperatively (SD)	n	3-months postoperatively (SD)	p-value
Intensity of activity					
% Sedentary	34	79.4 (7.8)	34	85.1 (7.2)	0.0008
% Light	34	18.5 (6.6)	34	13.4 (6.6)	0.0003
% MVPA	34	2.1 (2.5)	34	1.5 (1.3)	0.7648
Number steps/day	32	3191	32	2133	0.0047
Oxford Knee Score (OKS)	37	16	35	34	<0.0001
Hip					
Physical activity	n	Preoperatively (SD)	n	3-months postoperatively (SD)	p-value
Intensity of activity					
% Sedentary	32	78.6 (8.2)	32	78.8 (8.4)	0.6808
% Light	32	19.5 (7.2)	32	19.0 (7.3)	0.8959
% MVPA	32	2.0 (1.7)	32	2.2 (2.1)	0.5371
Number steps/day	32	3460	32	3559	0.6946
Oxford Hip Score (OHS)	33	16	33	37	<0.0001

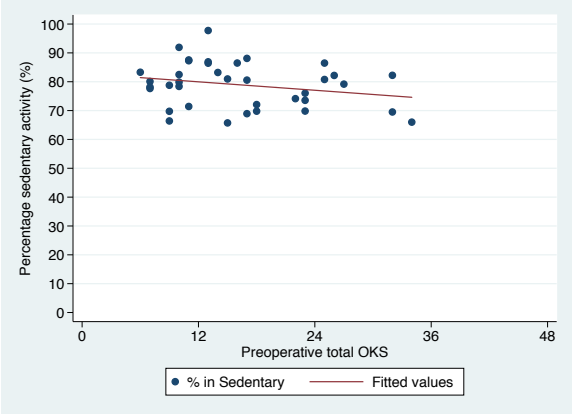
SD: Standard Deviation; MVPA: Moderate to vigorous physical activity

7.3.3 Is there a correlation between the OKS/OHS and accelerometer outcome measures?

Figures 5 and 6 are scatter plots of percentage time in sedentary activity and total OKS and OHS respectively, preoperatively and at three months postoperatively. They demonstrate that there is no significant correlation. Graphs 7 and 8 show a similar picture at the three month postoperative period for both TKA and THA patients respectively.

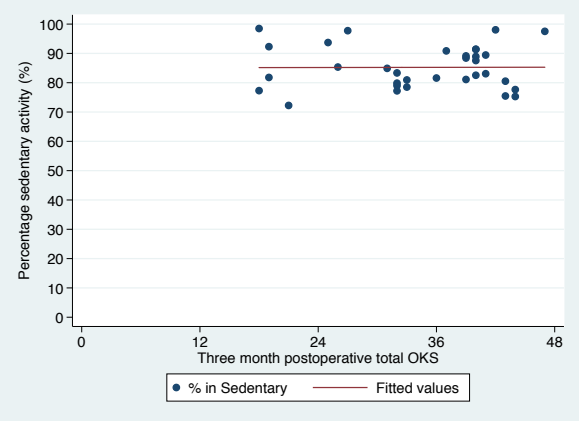
Figure 5: Scatter plot of showing percentage time in sedentary activity and preoperative total Oxford Knee Score (OKS) in TKA patients

(a) Preoperative



Spearman correlation: -0.17

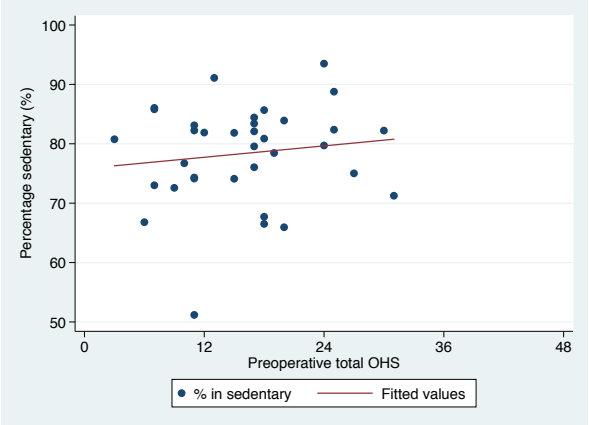
(b) Three-month post-operative



Spearman correlation: -0.04

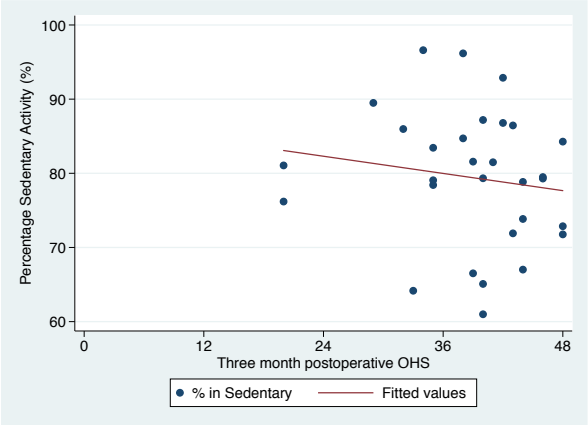
Figure 6: Scatter plot of percentage time in sedentary activity and preoperative total Oxford Hip Score (OHS) in THA patients

(a) Preoperative



Spearman correlation: 0.1027

(b) Three-month post-operative



Spearman correlation: -0.2113

A sub-group analysis was also undertaken to assess if there was a correlation between the constituent parts of the OKS and OHS and accelerometer data. Tables 4 and 5 shows the 12-item component questions of the OKS and OHS respectively and degree pf correlation with objective PA data from the AG3X. The OKS questions 4, 7, 8, 11 and 12 appeared to have a weak to moderate correlation with the number of steps taken per day and percentage time spent in sedentary activity. The OHS questions 5, 6 and 8 also had weak to moderate correlations to PA data. For both knee and hip patients, the question on walking before pain becomes severe (Q4 in OKS and Q6 in OHS) and the question on shopping (Q11 in OKS and Q5 in OHS) appeared to correlate moderately well with PA objective data.

Table 5: Correlation with preoperative Oxford Knee score constituent questions and accelerometer data (n=37)

	Spearman correlation (ρ)			
OKS questions	**No steps/day**	**p-value**	**% sedentary activity**	**p-value**
Q1: How would you describe the pain you usually have from your knee?	-0.0426	0.8025	-0.0145	0.9323
Q2: Have you had any trouble with washing and drying yourself (all over) because of your knee?	-0.0614	0.7179	0.1018	0.5488
Q3: Have you had any trouble getting in and out of a car or using public transport because of your knee?	0.0793	0.6408	-0.0399	0.8145
Q4: For how long have you been able to walk before pain from your knee becomes severe?	**0.3025**	0.0688	**-0.3470**	**0.0354**
Q5: After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your knee?	0.1038	0.5410	-0.1074	0.5271
Q6: Have you been limping when walking, because of your knee?	-0.0382	0.8225	-0.0592	0.7277
Q7: Could you kneel down and get up again afterwards?	**0.4703**	**0.0033**	**-0.4419**	**0.0062**
Q8: Have you been troubled by pain from your knee in bed at night?	-0.2708	0.1050	**0.3610**	**0.0282**
Q9: How much has pain from your knee interfered with your	0.2027	0.2288	-0.2677	0.1092

usual work (including housework)?				
Q10: Have you felt that your knee might suddenly 'give way' or let you down?	0.0696	0.6822	-0.2369	0.1580
Q11: Could you do the household shopping on your own?	**0.4057**	**0.0127**	-0.2938	0.0776
Q12: Could you walk down one flight of stairs?	**0.4644**	**0.0038**	**-0.4640**	**0.0038**

Figures in bold signifies where there is a significant (p-value <0.05) and weak (ρ=0.2-0.39) or moderate correlation (ρ=0.4-0.59)

Table 6: Correlation with preoperative Oxford Hip Score constituent questions and accelerometer data (n=36)

	Spearman correlation (ρ)			
OHS questions	**No steps/day**	**p-value**	**% sedentary activity**	**p-value**
Q1: How would you describe the pain you usually had from your hip?	0.0038	0.9822	0.2438	0.1519
Q2: Have you had any trouble with washing and drying yourself (all over) because of your hip?	0.1727	0.3137	0.0328	0.8496
Q3: Have you had any trouble getting in and out of a car or using public transport because of your hip?	0.0348	0.8403	0.0365	0.8326

Q4: Have you been able to put on a pair of socks, stockings or tights?	0.2482	0.1445	-0.0362	0.8341
Q5: Could you do the household shopping on your own?	**0.4149**	**0.0119**	-0.2091	0.2211
Q6: For how long have you been able to walk before pain from your hip becomes severe?	**0.4001**	**0.0156**	-0.1586	0.3557
Q7: Have you been able to climb a flight of stairs?	0.1648	0.3367	0.1239	0.4714
Q8: After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your hip?	-0.0761	0.6640	**0.3517**	**0.0383**
Q9: Have you been limping when walking, because of your hip?	-0.1941	0.2639	0.2690	0.1182
Q10: Have you had any sudden, severe pain - 'shooting', 'stabbing' or 'spasms' - from the affected hip?	0.0379	0.8289	0.0358	0.8381
Q11: How much has pain from your hip interfered with your usual work?	-0.0133	0.9398	0.2472	0.1522
Q12: Have you been troubled by pain from your hip in bed at night?	-0.1620	0.3524	0.1070	0.5406

Figures in bold signifies where there is a significant (p-value <0.05) and weak (ρ=0.2-0.39) or moderate correlation (ρ=0.4-0.59)

7.3.4 Practical aspects of accelerometer usage

In terms of acceptability and ease of use, all patients except one found the AG3X easy to use and comfortable to wear. One patient had to stop wearing the accelerometer because the waist belt was too constrictive, but this patient had a BMI >45. There was a 10% loss of devices over a 12-month period as a result of being lost in the post and failure of participants to return their devices.

7.4 Discussion

Accelerometers provide a vast amount of data on physical activity including energy expenditure, the number of steps taken in a day or per minute as well as percentage amount of time spent in sedentary/light/moderate to vigorous physical activity. Sedentary activity was defined as <100 counts/min, light physical activity as100-2019 counts/min and moderate-to-vigorous physical activity (MVPA) as >2020 counts/min(Liu et al. 2016; Troiano et al. 2008). In our study, we decided to use number of steps per day and percentage time in sedentary activity as markers of PA. Both these variables show a very strong correlation with one each other and have been used in many other studies involving the use of accelerometers in hip and knee replacement subjects(Höll et al. 2018; Harding et al. 2013; Brandes et al. 2010; Jiang et al. 2015; Thewlis et al. 2019).

Although one of the main reasons for performing a hip or knee replacement is to improve the function and mobility of a patient, the results from this study suggest that activity levels remain unchanged or indeed worsen after primary THR or TKR. The patients who had knee replacements had a significant improvement in their OKS, with a mean improvement of 18 points whilst patients who had a hip replacement had a mean improvement in their OHS of 21 points. Based on these patient-reported outcome measures (PROMs) results, one might conclude that the surgery was a great success and in terms

of pain relief and self-reported functional improvement, this is true(Meding et al. 2011). However, in terms of physical activity levels, there is a discrepancy between the subjective PROMs, such as the Oxford hip and knee scores and objective activity data as captured by the accelerometers. Our results are similar to other studies which have also shown a similar lack of improvement in activity levels when accelerometer data has been analysed(Harding et al. 2013; Paxton 2015; Thewlis et al. 2019.; Vissers et al. 2010; Vissers et al. 2013). In their study of 63 patients (of whom 44 (70%) had complete accelerometer data), Harding et al. found no improvement in objective activity levels after hip and knee replacement despite significant improvements in patient reported subjective outcome questionnaires with 82% of their patients being sedentary before surgery and 83% remaining sedentary six months postoperatively(Harding et al. 2013). In our study, 79.4% of knee replacement patients were sedentary preoperatively and this increased significantly to 85% at three months postoperatively. In hip replacement patients, 78% of patients were sedentary pre- and at three months post-operatively. This would suggest that there may be a difference in how quickly patients take to recover between hip and knee patients with hip patients returning back to their pre-operative activity levels quicker than knee replacement patients. In a systematic review of physical activity after TKR, Paxton et al. concluded that accelerometry-based studies demonstrated that physical activity in knee replacement patients continued to remain the same or indeed decreases after surgery(Paxton 2015). The authors also suggested that patients tended to self-report higher levels of function compared to objectively measured levels of activity using accelerometry(Paxton 2015). In a prospective study of 51 patients who had a primary THR, the patients spent 19.5h/day either sleeping or were sedentary and postoperatively, this did not change significantly(Thewlis et al. 2019).

In terms of how well accelerometer data correlated with OKS and OHS, our results suggest that there is no correlation with the number of steps taken per day or time spent in sedentary activity and total OKS/OHS. Thus, despite a significant improvement in OKS and OHS from preoperative to postoperative time at three months, there is no correlation between the objective

accelerometer data and subjective OKS/OHS. This is reflected in the fact that PA worsened in the TKA patients and remained static in the THA patients. This finding supports that of a small study involving 46 patients who underwent minimally invasive THA, whereby there was no correlation between the accelerometer data (measuring steps per day) and the outcome questionnaires Western Ontario and McMaster Universities Arthritis Index (WOMAC) and Harris Hip Score (HHS) (Höll et al. 2018). Unlike our results, the researchers did find that at 3 months, there was a significant increase in the number of steps per day after THA. This may reflect the fact that this was minimally invasive surgery and the patients were younger (mean age of 63) and had a lower BMI (mean of 27), thus likely fitter and more active(Höll et al. 2018). In a small study of 19 patients undergoing TKA, the authors did not find any correlation between self-reported mobility and objectively measured mobility(Walker et al. 2002).

There were some limitations to this study. Firstly, the sample size was relatively small and therefore, it is difficult to make sweeping conclusions regarding PA after hip or knee replacements and the significance of any potential correlation between accelerometer data and OHS/OKS. Nevertheless, this study is still one of the largest in the literature and the results will help provide recommendations for the main POSt-Plus study. Secondly, the follow up time period was at three months and this may not be sufficient to show any improvements in PA as longer follow up time points (6 or 12 months) might be preferable(de Groot et al. 2008). As a result, one of the recommendations for the main POSt Plus study is that accelerometers should be used at the 6 and 12-month postoperative follow up timepoint. Thirdly, the patients were not asked to complete a physical activity diary and therefore, the accelerometers may not record certain types of activities such as cycling or swimming (as the accelerometers are not waterproofed). A diary would help provide a more complete assessment of a patient's overall PA lifestyle, especially when used in conjunction with a validate PA questionnaire such as the IPAQ. This is another recommendation for the main study. Fourthly, the AG3X were worn at the waist and whilst theoretically, should be able to differentiate between a sitting, standing and lying down position, there is some reservation as to how

accurate this data is. Thus, there may be an element of under or over reporting of PA outputs. Finally, the cut points used in this study were recommended in the ActiGraph manual, but these cut points were based on data using a cohort of healthy fit male volunteers and not patients with OA awaiting THR/TKR and therefore the cut points used to determine what activity is sedentary/light or MVPA may not be valid in this type of population. A study by Lopes et al. looking at activity cut points in older, obese patients with diabetes showed no difference compared to the cut points suggested in the ActiGraph manual so we are confident that this issue is not a significant limitation(Lopes et al. 2009).

7.5 Conclusion

Based on the findings from this study, the literature review would support the use of accelerometers in the main POSt Plus study as the combination of accelerometer data together with self-reported questionnaires would provide a more complete picture of PA compared to just accelerometers or questionnaires on their own. Based on accelerometer data, PA does not improve at three months post-operatively after TKA or THA. There was only a very weak correlation between the final OKS/OHS and accelerometer data although some of the functional questions that make up the OKS/OHS did showed moderate correlation with accelerometer data. Finally, we would recommend the use of accelerometers in the main study in conjunction with questionnaires and would suggest adding in a diary to record daily wear of the AG3X as well as adding in a specific physical activity questionnaire such as the IPAQ. Whilst accelerometry provides objective data on PA, questionnaires and patient diaries provide additional information such as validation of the wear times for the accelerometer, as well as capturing data on activities which are not easily recorded by the AG3X such as swimming or cycling. This pilot study has shown that whilst there is some correlation with some of the functional aspects of the OKS and OHS, accelerometry data provides a much more detailed measure of PA and may help explain why there is sometimes a discrepancy between good OKS and OHS scores after TKA and THA yet poor

satisfaction. Accelerometry data may provide data which helps mitigate against the ceiling effect of the Oxford scores. This study has shown that the use of accelerometers is safe, reliable and that most patients were able to wear them safely and reliably.

Chapter 8 Summary and future directions

The objectives of POSt-Pilot were to incorporate patient public involvement (PPI) in order to help with the design and conduct of the main study (POSt-Main), determine the feasibility of the study, explore ways of managing and mitigating against missing and incorrect data and explore the use of accelerometers. To this end, the previous chapters in this thesis have achieved the objectives of the pilot study.

Chapter 3 described how patients were instrumental in helping develop and shape the study, using patient focus groups as the vehicle to gather data from patients. The focus group provided evidence that patients felt that the research subject was important for patients, was successful in helping provide feedback on the use of questionnaires and accelerometers in the pilot study as well as providing reassurance that excessive use of questionnaires was not likely to be a significant issue. As a result of the focus groups, changes were made to the design of the questionnaires and conduct of the pilot study.

Chapter 4 provided information about the overall general feasibility of the pilot study and concluded that the study was feasible in terms of consent rates and retention. Furthermore, the potential for questionnaire burnout was not seen and patients tolerated wearing the accelerometers with only one patient having difficulty with the accelerometers. The main issues identified in the pilot was that recruitment rates were low and that an error was identified in the use of the TSK-11 questionnaire. The pilot demonstrated that additional research staff support was required to enable research team members to consent and recruit patients from multiple sites and to provide cover in the event that a member of the research team was away. Contacting patients by telephone before they attended pre-assessment clinic was also a good way of optimizing recruitment, ensuring potential participants had received the patient information sheet and were expecting to be met by a member of the POSt research team.

Chapter 5 analysed a method for managing the issue of using an incorrect questionnaire in the pilot, in this case, the TSK-11 questionnaire. The chapter provided a detailed plan for avoiding such an error in the main study and also explored options for mitigating against such an error if this were to occur again. Chapter 6 assessed the levels of missing data in the variables BMI and OHS/OKS and provided a detailed plan on how to minimize missing data in terms of how POSt was designed and conducted. The chapter also describes a data analysis plan to manage missing data using MI as a model.

Finally, chapter 7 reviewed the use of accelerometers and showed that there was a mismatch between a patient's subjective measure of PA and the accelerometer's objective measure of PA. Self-reported activity related questionnaires may report higher levels of PA than is measured using accelerometers. The chapter also identified one or two key outputs which were found to be good markers of PA, namely the number of steps taken daily and percentage time spent in various degrees of PA (sedentary to vigorous activity). The use of accelerometers was very well tolerated by patients in the pilot study and the data captured provided a different perspective on PA compared to that collected from questionnaires.

Going forwards, the pilot study has provided a framework for the main research study. The pilot study has shown that the research question is important to patients and the objectives are valid. The study protocol is appropriate, and the consent and retention rates are good and overall, the primary objective of the pilot was achieved, namely that the study is feasible. Recruitment rate is a concern but with additional resources in terms of recruiting more research team members and putting in place telephone reviews, it is hoped that the recruitment rate will pick up. The pilot identified an error in the use of an incorrect questionnaire (TSK-11) which if this had continued to be used in the main study, could have potentially been a serious issue as the data from the incorrect questionnaire may not useable in the analysis. This issue was explored and there is a process which will be followed to ensure that in future,

the correct questionnaires are used but if an error did occur regarding using an incorrect version of a questionnaire, then one option would be to try and use multiple imputation as a way of still using the available data. Missing data in any research study is always an issue that needs addressing. One of the objectives of the pilot study was to review the levels of missing data and develop an analysis plan for how to manage such missing data. The outcome of the pilot showed that by putting into place a number of processes in the design and conduct of the study, the levels of missing data could be minimized as much as possible. The use of multiple imputation was shown to be a potential method of handling missing data, at the unit level but also at the item level. Whilst outside the scope of this pilot study, in view of the small sample size, using multiple imputation by chained equations (MICE) and including a greater number of covariates in the predictive MI model, may help mitigate against missing data especially in the primary outcome measures of the Oxford hip and knee scores. The pilot also suggested that using the Oxford rules if only one or two items were missing from the 12-item score in order to still calculate a total score, seemed a reasonable strategy and this may even be applicable if 3 items were missing. To conclude, the POSt pilot has successfully proven that the study is feasible with the use of appropriate questionnaires and accelerometers and a valid data analysis plan has been created to handle incorrect or missing data.

Appendices

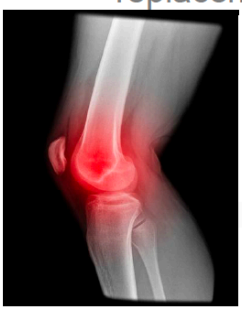
Appendix 1: Poster for clinics to help with patient recruitment

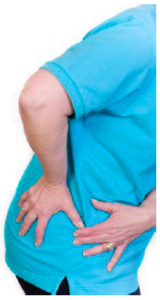
University Hospitals of Leicester NHS
NHS Trust

University of Leicester

POSt

***P**redicting **O**utcome **St**udy after hip and knee replacements*





Chief Investigator: Prof J Dias Principal Investigator: Mr J Palan

1. Do you suffer from osteoarthritis?
2. Are you about to have your first total hip or knee replacement?
3. Would you be interested in taking part in this clinical study?

This study involves the use of questionnaires and an activity monitor, to help understand why a few patients are dissatisfied after their first hip or knee replacement.

If your answer is YES to all three questions, please contact the research team (see contact details below) for further information and register your interest.

An invitation letter and patient information sheet will be mailed to you. A member of the research team will then contact you, and meet you when you arrive for your pre-admission clinic to discuss the study further.

This study has been given a Favourable Opinion by the Leicester Research Ethics Committee (Clinical study REC reference: 13/EM/0061) and is sponsored by UHL NHS Trust.

For further information:
Mrs Lynne Clarke, Senior Clerk
Department of Health Sciences
Orthopaedic Section
University of Leicester
Leicester General Hospital
Clinical Sciences Unit - Off Ward 11
Gwendolen Road
LE5 4PW
Tel: +44 (0)116 2588041
Email: lg39@leicester.ac.uk
http://www.hs.le.ac.uk

Poster (Version_03/1.8.13)

Appendix 2: Patient Information Sheet (PIS)

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

School of Medicine
Department of Health Sciences
Clinical Division of Surgery
Orthopaedic Surgery Section
Clinical Sciences Unit - Off Ward 11
The Leicester General Hospital
Gwendolen Road
Leicester LE5 4PW
UK

Head of Section
Professor J J Dias

T +44 (0)116 258 4702/8041/4706

Patient information sheet

Predicting Outcome Study (POSt) following hip and knee replacements

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

Who are we?

My name is Mr Jeya Palan and I am an orthopaedic specialist registrar working with the following members of the research team:

Professor J Dias (Professor of Orthopaedic Surgery); Professor P Watson (Professor of Pain Management); Mr C Esler (Consultant Orthopaedic Surgeon); Prof J Thompson (Professor of Medical Statistics); Dr N Taub (Lecturer in Medical Statistics); Prof G Macfarlane (Professor of Medical Epidemiology)

What is the purpose of the study?

The primary aim of this study is to identify the risk factors that may predict which patients are likely to be dissatisfied after their hip or knee replacement.

The study will last about 3 years from the time patients are invited to take part in the study to the time the results are analysed and published.

Although hip and knee replacement operations are extremely successful in relieving pain and improving patients' function and quality of life, up to 19% of patients worldwide (about 20,000 patients every year in the UK) are not satisfied following their hip or knee replacements. This is despite many advances in medical technology and surgical techniques over the last 20 or so years. We still do not know why this is the case although certain factors are likely to make patients dissatisfied. These factors include both physical and psychological factors. It is very important to identify these factors so that such patients can be given targeted and patient-specific treatment before surgery to help improve their outcome after surgery.

Why have I been chosen?

You have been chosen because you are about to have your first hip or knee replacement and have a diagnosis of primary osteoarthritis of the hip or knee joint. You also have a good command of the English language and can understand and complete the questionnaires without the need for a language interpreter*

*It is not our intention to exclude patients from participating in this study but for practical and financial reasons, we do not have the resources to translate the questionnaires into other languages other than English. Furthermore, these questionnaires have been designed for use with the English language and thus translating the questionnaires into other languages may affect the reliability and validity of the responses.

In total, up to 400 patients will be invited to take part in this study, with 100 patients being invited to take part initially.

Can I decide not to take part in the study?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. This will not affect the standard of care you receive.

What will happen to me if I take part?

This study does NOT involve the use of any new treatments or medication. It is a study designed to observe how patients feel and function both before and after their hip or knee replacement surgery.

This study will involve the use of questionnaires, designed to help assess how patients' perceive pain, their level of anxiety, depression and other psychological factors. We will also be using specific questionnaires designed to assess general health and quality of life and more specific ones for hip and knee arthritis. In total, there will be 9 different questionnaires involved for each patient to complete, although not all questionnaires will need to be completed at once.

This study will also use a special device that monitors physical activity in the home environment. This is called an accelerometer and is a lightweight device (like a small watch) that can be worn on the wrist or around the waist (as discretely as possible). It is battery operated with a charging dock and records activity such as the number of steps taken and the types of activity (for example, going up or down stairs) involved. The device will be worn for up to 10 hours a day whilst the patient is active and mobile, for 1 week before the operation and 1 week after the operation. The device can be removed for overnight charging whilst the patient sleeps.

A flowchart is provided to show the timeline for when the questionnaires will need to be completed for each patient and when the activity monitors will need to be worn.

Are there any additional risks or burdens involved?

There are no additional risks to your health (other than the risks of having surgery and an anaesthetic, as part of having your hip or knee replacement). The main burden will be on your time, in terms of completing the questionnaires. The activity monitors are very small and comfortable to wear.

What if something goes wrong?

We do not anticipate any new complications arising as a result of participating in this study. However, as a patient undergoing either a total hip or knee replacement, you will still be at risk of the potential complications arising as a result of having an operation. Should such a complication arise, you will be treated appropriately but this would not preclude you from continuing to be a part of the study. However, if you so wish, you are free to withdraw from this study at any time.

What will I gain from being involved?

It is hoped that the process of taking part in the study may help clarify any concerns patients may have regarding having a hip or knee replacement operation. It may provide patients with a better understanding of their hip or knee replacement.

Where will the study take place?

The study is based at the Leicester General Hospital in conjunction with the University of Leicester.

Will the information provided be confidential and will I remain anonymous?

There will be complete and total confidentiality regarding the use of any information provided by the patients taking part in the study. This information will not be shared with any third party commercial groups. The information provided will be stored on an approved NHS encrypted memory stick and

computer and all necessary precautions regarding patient confidentiality will be taken.

Personal data will be stored on University of Leicester computers based at Leicester General Hospital and may be shared with authorised colleagues at the University of Leicester, regulatory bodies and University Hospitals Leicester NHS Trust. This data will be kept for a maximum of 36 months following the end of the study and may be used to re-contact participants to invite them to take part in future related research studies.

How will the information provided by the study be used?

The results of this study will hopefully provide us with a method of screening patients who are at higher risk than normal of being dissatisfied after their hip or knee replacement. This would then allow us to try different therapies with the eventual aim of improving the overall patient experience following hip and knee replacement surgery. The findings from this study will be presented at conferences and published in peer reviewed medical journals. It will also form the main part of a PhD study. At no point will any patients be identified in any presentations or papers.

Who has funded this study?

This study has been funded by the Foxtrot Charitable Trust and it is hoped, will also be funded by Arthritis Research UK. No pharmaceutical or industry funding is involved in this study.

Who has reviewed this study?

This study has been reviewed by the NRES Committee East Midlands (Leicester) on the 1st February 2013 and has been given a favourable opinion.

Thank you.

Any further questions, please contact:

Mr Jeya Palan
Trauma and Orthopaedic Specialist Registrar
Department of Trauma and Orthopaedics
Department of Health Sciences
Room 6
New Academic Unit
Leicester General Hospital
Gwendolen Road
Leicester LE5 4PW
Email: jp314@le.ac.uk
Tel: 0116 258 4899

Appendix 3: Patient consent form for POSt Pilot

University Hospitals of Leicester NHS
NHS Trust

Centre Number: University Hospital Leicester NHS Trust (Leicester General Hospital)

Study Number: 11189

Patient Identification Number for this trial:

CONSENT FORM

Title of Project: **Predicting Outcome following total Hip and Knee replacement Study (POSt)**

Name of Researcher: **Professor Joe Dias (Chief Investigator); Mr Jeya Palan (Principal Investigator)**

Please initial all boxes

1. I confirm that I have read and understand the Patient Information Sheet (PIS) dated 15.2.13 (version 5) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that relevant sections of my medical notes (including personal data) and data collected during the study, may be looked at by individuals from the University of Leicester, from regulatory authorities or from the NHS Trust (University Hospitals Leicester), where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐

4. I understand and agree for my personal data to be stored and shared by authorised personnel at the University of Leicester. ☐

5. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of Person taking consent	Date	Signature

Consent form date of issue: 15.2.13
Consent form version number: 06

Appendix 4: Actigraph GT3X Patient information sheet

University Hospitals of Leicester NHS
NHS Trust

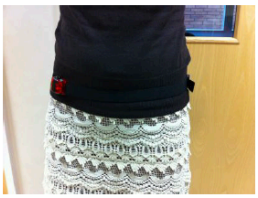
POSt

Predicting Outcomes Study after total hip and knee replacement

Instructions on how to wear your activity monitor

This is a small activity monitor that records general movement and allows us to get a better idea of your overall activity level. We will NOT be able to tell what kind of specific activity is happening. It is very important that the activity monitor is worn properly. If it is not worn properly, we may have to ask you to wear the activity monitor again. Please follow the instructions below to ensure that the activity monitor is worn correctly.

Picture A



Picture B

1. Wear the activity monitor attached to the belt around your waist just above the RIGHT HIP BONE (as shown in the Picture A).
2. Please check that the star on the monitor faces outwards and that the number and black button on the monitor is facing upwards (as shown in Picture B).
3. Wear the monitor snug against your body. If you have to, you can adjust the belt by pulling on the end of the strap to make it tighter. You can loosen the belt by pushing more of the strap through the loop.
4. Put on the monitor first thing in the morning either just after you wake up or have your bath/shower.
5. Please take the monitor off last thing at night, before you go to sleep.
6. Please wear the monitor for **7 days** continuously.
7. Please wear the monitor for a minimum of **10 hours/day**.
8. Please take off the monitor when showering, bathing or swimming and don't forget to wear the Activity monitor again.
9. Don't let anyone else wear the monitor.
10. Please look after the monitor and post the monitor back as soon as you have worn it for 7 days, in the Jiffy Bag provided (which is prepaid and self addressed already).

Appendix 5: Oxford Hip Score

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Oxford Hip Score

The following 12 questions look at how your hip has been in the **last four weeks.**

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

How would you describe the pain you usually have from your hip?

None	Very Mild	Mild	Moderate	Severe
☐	☐	☐	☐	☐

Have you had any trouble with washing and drying yourself (all over) BECAUSE OF YOUR HIP?

No Trouble	Very Little Trouble	Moderate Trouble	Extreme Difficulty	Impossible to Do
☐	☐	☐	☐	☐

Have you had any trouble getting in and out of a car or using public transport BECAUSE OF YOUR HIP?

No Trouble	Very Little Trouble	Moderate Trouble	Extreme Difficulty	Impossible to Do
☐	☐	☐	☐	☐

Have you been able to put on a pair of socks, stockings or tights?

Yes, easily	With Little Difficulty	With Moderate Difficulty	With Extreme Difficulty	No, impossible
☐	☐	☐	☐	☐

Could you do the household shopping ON YOUR OWN?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

How long have you been able to walk before PAIN FROM YOUR HIP becomes SEVERE? (with or with a stick)

More than 30 minutes	16-30 Minutes	5 to 15 Minutes	Around the House Only	Not at all
☐	☐	☐	☐	☐

Have you been able to climb ONE flight of stairs?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

After a meal (sat at a table), how painful has it been to stand up from a chair BECAUSE OF YOUR HIP?

Not at all	Slightly	Moderately	Very	Unbearable
☐	☐	☐	☐	☐

Have you been limping when walking BECAUSE OF YOUR HIP?

Rarely/Never	Sometimes, or just at first	Often, not just at first	Most of the time	All of the time
☐	☐	☐	☐	☐

Have you had any sudden SEVERE pain - 'shooting', 'stabbing' or 'spasms' from the AFFECTED HIP?

No	Only 1 or 2 days	Some days	Most days	Every day
☐	☐	☐	☐	☐

How much has PAIN FROM YOUR HIP interfered with your usual work (including housework)?

Not at all	A little bit	Moderately	Greatly	Totally
☐	☐	☐	☐	☐

Have you been troubled by PAIN FROM YOUR HIP in bed at night?

No	Only 1 or 2 nights	Some nights	Most nights	Every night
☐	☐	☐	☐	☐

Office Use Only

Patient

Time Code

Appendix 6: Oxford Knee Score

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Oxford Knee Score

The following 12 questions look at how your knee has been in the **last four weeks.**

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

How would you describe the pain you usually have from your knee?

None	Very Mild	Mild	Moderate	Severe
☐	☐	☐	☐	☐

Have you had any trouble with washing and drying yourself (all over) BECAUSE OF YOUR KNEE?

No Trouble	Very Little Trouble	Moderate Trouble	Extreme Difficulty	Impossible to Do
☐	☐	☐	☐	☐

Have you had any trouble getting in and out of a car or using public transport BECAUSE OF YOUR KNEE?

No Trouble	Very Little Trouble	Moderate Trouble	Extreme Difficulty	Impossible to Do
☐	☐	☐	☐	☐

How long have you been able to walk before PAIN FROM YOUR KNEE becomes SEVERE? (with or without a stick)

More than 30 Minutes	16 - 30 Minutes	5 - 15 Minutes	Around the House Only	Not at All
☐	☐	☐	☐	☐

After a meal (sat at a table), how painful has it been to stand up from a chair BECAUSE OF YOUR KNEE?

Not at all	Slightly	Moderately	Very	Unbearable
☐	☐	☐	☐	☐

Have you been limping when walking BECAUSE OF YOUR KNEE?

Rarely/Never	Sometimes, or just at first	Often, not just at first	Most of the time	All of the time
☐	☐	☐	☐	☐

Could you kneel down and get up again afterwards?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

During the last 4 weeks... Mark **ONE** box for **EVERY** answer with a ☒

Have you been troubled by PAIN FROM YOUR KNEE in bed at night?

No	Only 1 or 2 nights	Some nights	Most nights	Every night
☐	☐	☐	☐	☐

How much has PAIN FROM YOUR KNEE interfered with your usual work (including housework)?

Not at all	A little bit	Moderately	Greatly	Totally
☐	☐	☐	☐	☐

Have you felt that your knee might suddenly 'give way' or let you down?

Rarely/Never	Sometimes, or just at first	Often, not just at first	Most of the time	All of the time
☐	☐	☐	☐	☐

Could you do the household shopping ON YOUR OWN?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

Could you walk down ONE flight of stairs?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐	☐	☐	☐	☐

Office Use Only

Patient	Time Code

Appendix 7: Brief Pain Inventory

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

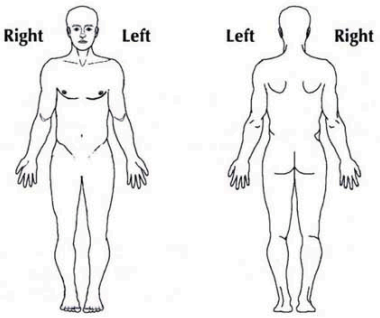
Brief Pain Inventory

First Name | Surname | Date D D / M M / Y Y Y Y | Time H H : M M

1) **Throughout our lives, most of us have had pain from time to time (such as minor headaches, sprains, and toothaches). Have you had pain other than these everday kinds of pain today?**

☐ Yes ☐ No

2) **On the diagram, mark with an X, the areas where you feel pain. Circle the X that hurts the most.**



3) **Please rate your pain by ticking the number that best describes your pain at its WORST in the last 24 hours:**

0 1 2 3 4 5 6 7 8 9 10

No Pain — Pain as bad as you can imagine

4) **Please rate your pain by ticking the number that best describes your pain at its LEAST in the last 24 hours:**

0 1 2 3 4 5 6 7 8 9 10

No Pain — Pain as bad as you can imagine

5) **Please rate your pain by ticking the number that best describes your AVERAGE pain:**

0 1 2 3 4 5 6 7 8 9 10

No Pain — Pain as bad as you can imagine

6) **Please rate your pain by ticking the number that best describes your pain RIGHT NOW:**

0 1 2 3 4 5 6 7 8 9 10

No Pain — Pain as bad as you can imagine

7) **What treatments or medications are you receiving for your pain?**

8) **In the last 24 hours, how much relief have pain treatments or medications provided? Please tick the one percentage that shows how much RELIEF you have received:**

0 10 20 30 40 50 60 70 80 90 100

No Relief — Complete Relief

9) **Tick the one number that describes how, during the past 24 hours, pain has interfered with your:**

a) **General Activity**

0 1 2 3 4 5 6 7 8 9 10

Does not interfere — Completely interferes

b) **Mood**

0 1 2 3 4 5 6 7 8 9 10

Does not interfere — Completely interferes

c) **Walking ability**

0 1 2 3 4 5 6 7 8 9 10

Does not interfere — Completely interferes

d) **Normal work (i.e. work outside the home and housework**

0 1 2 3 4 5 6 7 8 9 10

Does not interfere — Completely interferes

e) **Relations with other people**

0 1 2 3 4 5 6 7 8 9 10

Does not interfere — Completely interferes

f) **Sleep**

0 1 2 3 4 5 6 7 8 9 10

Does not interfere — Completely interferes

g) **Enjoyment of life**

0 1 2 3 4 5 6 7 8 9 10

Does not interfere — Completely interferes

In addition to completing the Brief Pain Inventory, to help your doctor better manage your pain, please tell us:

What does the pain feel like? Tick the words that decribe your pain

☐ aching	☐ exhausting	☐ squeezing	☐ tender	☐ dull	☐ pricking	☐ miserable
☐ stabbing	☐ nagging	☐ throbbing	☐ tiring	☐ cramping	☐ burning	☐ radiating
☐ sharp	☐ unbearable	☐ gnawing	☐ numb	☐ shooting	☐ penetrating	☐ deep

How long have you had this pain?

☐ less than a week ☐ 1 to 2 weeks ☐ 2 to 4 weeks ☐ more than a month

What kind of things make your pain feel better (i.e. heat, medicine, rest)?

What kind of things make your pain feel worse (i.e. walking, standing, lifting)?

Do you have any other symptoms?

☐ nausea	☐ vomiting	☐ constipation	☐ diarrhea	☐ lack of appetite	☐ indigestion
☐ difficulty sleeping	☐ feeling drowsy	☐ nightmares	☐ dizziness	☐ tiredness	☐ itching
☐ urinary problems	☐ sweating	☐ weakness	☐ headaches		

Talking About Your Pain

It's important to remember that each person's pain is different. The pain that you experience cannot be compared to another person's pain. ONLY YOU know how and when you hurt, and how the pain affects your life.

It is important to describe what you are feeling to those who are trained to help you. Don't be embarrassed to talk to your doctor, nurse or pharmacist. They need to know as much as possible about your pain in order to develop the best plan to control it. The questions on this form can help you describe your pain.

Why is Pain Relief So Important?

Proper treatment for pain is not only a matter of comfort. Unrelieved pain can lead to nausea, loss of sleep, depression, loss of appetite, weakness, and other problems. Pain can also affect your life at home and at work. Relieving your pain means that you can continue to do the day-to-day things that are important to you.

Most Pain Can Be Controlled

It is important to know that most pain CAN be relieved. Your doctor will work with you to find the treatment that may be best for your pain.

The key to effective pain control is to take the RIGHT AMOUNT, of the RIGHT MEDICINE, at the RIGHT TIME. You should take your pain medicine on a regular schedule, as your doctor, nurse or pharmacist tells you. Don't wait until the pain becomes severe. Pain is easier to control when it is mild than when it has reached full force.

If your pain medicine wears off too soon, is not relieving the pain, or causes problems with side effects, you should call your doctor because you may need to have your treatment plan changed.

Comments: *Write down any questions or information you need to share with your doctor, nurse or pharmacist about your pain*

Office Use Only

Patient **Time Code**

Appendix 8: Hospital Anxiety and Depression Scale

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Hospital Anxiety and Depression Scale (HADS)

Patients are asked to choose one response from the four given for each statement. You should give an immediate response and shouldn't think too long about your answers. Please answer how it currently describes your feelings.

A I feel tense or 'wound up'

- ☐ Most of the time (3)
- ☐ A lot of the time (2)
- ☐ From time to time, occasionally (1)
- ☐ Not at all (0)

D I still enjoy the things I used to enjoy

- ☐ Definitely as much (0)
- ☐ Not quite so much (1)
- ☐ Only a little (2)
- ☐ Hardly at all (3)

A I get a sort of frightened feeling as if something awful is about to happen

- ☐ Very definitely and quite badly (3)
- ☐ Yes, but not too badly (2)
- ☐ A little, but it doesn't worry me (1)
- ☐ Not at all (0)

D I can laugh and see the funny side of things

- ☐ As much as I always could (0)
- ☐ Not quite so much now (1)
- ☐ Definitely not so much now (2)
- ☐ Not at all (3)

A Worrying thoughts go through my mind

- ☐ A great deal of the time (3)
- ☐ A lot of the time (2)
- ☐ From time to time, but not too often (1)
- ☐ Only occasionally (0)

D I feel cheerful

- ☐ Not at all (3)
- ☐ Not often (2)
- ☐ Sometimes (1)
- ☐ Most of the time (0)

A I can sit at ease and feel relaxed

- ☐ Definitely (0)
- ☐ Usually (1)
- ☐ Not Often (2)
- ☐ Not at all (3)

D I feel as if I am slowed down

- ☐ Nearly all the time (3)
- ☐ Very often (2)
- ☐ Sometimes (1)
- ☐ Not at all (0)

A I get a sort of frightened feeling like 'butterflies' in my stomach

- ☐ Not at all (0)
- ☐ Occasionally (1)
- ☐ Quite Often (2)
- ☐ Very Often (3)

D I have lost interest in my appearance

- ☐ Definitely (3)
- ☐ I don't take as much care as I should (2)
- ☐ I may not take quite as much care (1)
- ☐ I take just as much care as ever (0)

A I feel restless as I have to be on the move

- ☐ Very much indeed (3)
- ☐ Quite a lot (2)
- ☐ Not very much (1)
- ☐ Not at all (0)

D I look forward with enjoyment to things

- ☐ As much as I ever did (0)
- ☐ Rather less than I used to (1)
- ☐ Definitely less than I used to (2)
- ☐ Hardly at all (3)

A I get sudden feelings of panic

- ☐ Very often indeed (3)
- ☐ Quite often (2)
- ☐ Not very often (1)
- ☐ Not at all (0)

D I can enjoy a good book or radio or TV programme

- ☐ Definitely as much (0)
- ☐ Not quite so much (1)
- ☐ Only a little (2)
- ☐ Hardly at all (3)

Appendix 9: Self Efficacy for Rehabilitation

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Self Efficacy for Rehabilitation Scale (SER)

Reading the below statements, please tick the relevant box, rating from 'Cannot Do It' to 'Certain I Can Do It'

During my rehabilitation I believe I can do...	Cannot Do It 1	2	3	4	5	6	7	8	9	Certain I Can Do It 10
Therapy that requires me to stretch my leg	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Therapy that requires me to lift my leg	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Therapy that requires me to bend my leg	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Therapy that requires me to stand	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Therapy that requires me to walk	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
All of my therapy exercises	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
My therapy every day that it is scheduled	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
The exercises my therapists say I should do, even if I don't understand how it helps me	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
My therapy no matter how I feel emotionally	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
My therapy no matter how tired I may feel	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
My therapy even though I may already have other complicating illnesses	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
My therapy regardless of the amount of pain I am feeling	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Appendix 10: Pain Catastrophizing Score

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Pain Catastropising Score (PCS)

Read the statements and give your answer, by ticking the relevant box, rating from 'all the time' to 'not at all'

	All the time				Not at all
	4	3	2	1	0
I keep thinking about how badly I want the pain to stop	☐	☐	☐	☐	☐
I anxiouly want the pain to go away	☐	☐	☐	☐	☐
I can't seem to keep it out of my mind	☐	☐	☐	☐	☐
I keep thinking about how much it hurts	☐	☐	☐	☐	☐
I wonder whether something serious may happen	☐	☐	☐	☐	☐
I become afraid that the pain may get worse	☐	☐	☐	☐	☐
I think of other painful experiences	☐	☐	☐	☐	☐
I feel I can't go on	☐	☐	☐	☐	☐
It's terrible and I think it's never going to get any better	☐	☐	☐	☐	☐
I worry all the time about whether the pain will end	☐	☐	☐	☐	☐
It's awful and I feel that it overwhelms me	☐	☐	☐	☐	☐
I feel I can't stand it any more	☐	☐	☐	☐	☐
There is nothing I can do to reduce the intensity of the pain	☐	☐	☐	☐	☐

Appendix 11: EQ5D

EuroQol EQ-5D

(EuroQol used with permission)

By placing a cross (☒) in one box in each group below, please indicate which statement best describes your own health state **TODAY.**

Q1 **Mobility**

- I have **no** problem walking about ☐
- I have **some** problem walking about ☐
- I am **confined to bed** ☐

Q2 **Self Care**

- I have **no** problem with self care ☐
- I have **some** problem with self care ☐
- I am **unable to wash or dress myself** ☐

Q3 **Usual Activities**
(e.g. work, housework, family or leisure activities)

- I have **no** problem with my normal activities ☐
- I have **some** problem with my normal activities ☐
- I am **unable to perform my normal activities** ☐

Q4 **Pain/Discomfort**

- I have **no** pain or discomfort ☐
- I have **moderate** pain or discomfort ☐
- I have **extreme** pain or discomfort ☐

Q5 **Anxiety/Depression**

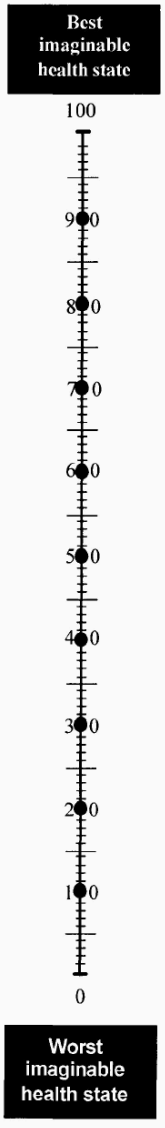
- I am **not** anxious or depressed ☐
- I am **moderately** anxious or depressed ☐
- I am **extremely** anxious or depressed ☐

Your own health state today

Think about how good or bad your own health is today.

This scale may help. The best health you can imagine is marked 100, and the worst health you can imagine is marked 0.

Please indicated, with a cross on the scale, how good or bad your own health is today.



Appendix 12: Self-Administered Patient Satisfaction Questionnaire

Self Administered Patient Satisfaction (SAPS) Questionnaire

How satisfied are you with the results of your surgery:

	Very Satisfied	Somewhat Satisfied	Somewhat Dissatisfied	Dissatisfied
In general	☐	☐	☐	☐
For improving your pain	☐	☐	☐	☐
For improving your ability to do home or garden work	☐	☐	☐	☐
For improving your ability to do recreational activities	☐	☐	☐	☐

Appendix 13: HSS Knee Surgery Expectations Survey

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Hospital for Special Surgery - Knee Surgery Expectations Survey

How much relief or improvement do you expect in the following areas as a result of your knee replacement?

	Back to Normal	**Not back to normal but...**			**I do not have this expectation** *or*
	Complete improvement	A lot of improvement	A moderate amount of improvement	A little improvement	This does not apply to me
Relief of pain	☐	☐	☐	☐	☐
Improve ability to walk - short distance (indoors, down the street)	☐	☐	☐	☐	☐
Improve ability to walk - medium distance (take a walk up to 1 mile)	☐	☐	☐	☐	☐
Improve ability to walk - long distance (more than 1 mile)	☐	☐	☐	☐	☐
Remove the need for a cane, crutch or walker	☐	☐	☐	☐	☐
Make knee or leg straight	☐	☐	☐	☐	☐
Improve ability to go up stairs	☐	☐	☐	☐	☐
Improve ability to go down stairs	☐	☐	☐	☐	☐
Improve ability to kneel	☐	☐	☐	☐	☐
Improve ability to squat	☐	☐	☐	☐	☐
Improve ability to use public transportation or drive	☐	☐	☐	☐	☐
Be employed for monetary reimbursement	☐	☐	☐	☐	☐
Improve ability to participate in recreation (i.e. dancing, pleasure travel)	☐	☐	☐	☐	☐
Improve ability to perform daily activites (i.e. household chores, daily routine)	☐	☐	☐	☐	☐
Improve ability to exercise or participate in sports	☐	☐	☐	☐	☐
Improve ability to change position (i.e. go from sitting to standing or vice versa)	☐	☐	☐	☐	☐
Improve ability to interact with others (i.e. take care of someone, play with children)	☐	☐	☐	☐	☐
Improve sexual activity	☐	☐	☐	☐	☐
improve psychological well-being	☐	☐	☐	☐	☐

Appendix 14: HSS Hip Surgery Expectations Survey

University Hospitals of Leicester NHS
NHS Trust

University of Leicester

Hospital for Special Surgery - Hip Surgery Expectations Survey

How much relief or improvement do you expect in the following areas as a result of your hip replacement?

	Back to Normal	**Not back to normal but...**			**I do not have this expectation** *or*
	Complete improvement	A lot of improvement	A moderate amount of improvement	A little improvement	This does not apply to me
Relief of daytime pain	☐	☐	☐	☐	☐
Relief of pain that interferes with sleep	☐	☐	☐	☐	☐
Improve ability to walk	☐	☐	☐	☐	☐
Improve ability to stand	☐	☐	☐	☐	☐
Get rid of limp	☐	☐	☐	☐	☐
Remove the need for a cane or other assistive device	☐	☐	☐	☐	☐
Improve ability to climb stairs	☐	☐	☐	☐	☐
Improve ability to get in or out of a bed, chair or car	☐	☐	☐	☐	☐
Improve ability to perform daily activites around the home (i.e. housework, gardening)	☐	☐	☐	☐	☐
Improve ability to perform daily activites away from the home (i.e. shopping, volunteer work)	☐	☐	☐	☐	☐
Eliminate the need for medications	☐	☐	☐	☐	☐
Be employed for monetary reimbursement	☐	☐	☐	☐	☐
Improve sexual activity	☐	☐	☐	☐	☐
Improve ability to exercise or participate in sports	☐	☐	☐	☐	☐
Improve ability to participate in social activities or recreation	☐	☐	☐	☐	☐
Improve ability to put on shoes and socks	☐	☐	☐	☐	☐
Improve ability to cut toenails	☐	☐	☐	☐	☐
Improve psychological well-being	☐	☐	☐	☐	☐

Appendix 15: Tampa Scale of Kinsesiophobia-11 (Correct version)

Tampa Scale of Kinesiophobia-11 (TSK-11) Correct

Read the statements and put a tick in the box which reflects your agreement with the statement:

	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
I'm afraid that I might injure myself if I exercise	☐	☐	☐	☐
If I were to overcome it, my pain would increase	☐	☐	☐	☐
My body is telling me I have something dangerously wrong	☐	☐	☐	☐
People aren't taking my medical condition seriously enough	☐	☐	☐	☐
My accident has put my body at risk for the rest of my life	☐	☐	☐	☐
Pain always means I have injured my body	☐	☐	☐	☐
Simply being careful that I do not make any unnecessary movements is the safest thing I can do to prevent my pain from worsening	☐	☐	☐	☐
I wouldn't have this much pain if there wasn't something potentially dangerous going on in my body	☐	☐	☐	☐
Pain lets me know when to stop exercising so that I don't injure myself	☐	☐	☐	☐
I can't do all the things normal people do because it's too easy for me to get injured	☐	☐	☐	☐
No one should have to exercise when s/he is in pain	☐	☐	☐	☐

Bibliography

1. Aaron, R.K., 2006. Arthroscopic Débridement for Osteoarthritis of the Knee. *The Journal of Bone and Joint Surgery (American)*, 88(5), pp.936–943.
2. Ackerman, I.N and Bennell, K.L. 2004. Does pre-operative physiotherapy improve outcomes from lower limb joint replacement surgery? A systematic review. *Australian Journal of Physiotherapy* 50: pp. 25–30
3. Ali, A. et al. 2014. Dissatisfied patients after total knee arthroplasty: a registry study involving 114 patients with 8-13 years of followup. *Acta Orthopaedica*, 85(3), pp.229–233.
4. Allami, M.K. et al., 2006. Outcome of Charnley total hip replacement across a single health region in England - The results at ten years from a regional arthroplasty register. *Journal of Bone & Joint Surgery, British Volume*, 88B(10), pp.1293–1298.
5. Allen, K.D. & Golightly, Y.M., 2015. Epidemiology of osteoarthritis: State of the evidence. *Current Opinion in Rheumatology*, 27(3), pp.276–283.
6. Allison, P.D., 2000. Multiple Imputation for Missing Data. *Sociological Methods & Research*.
7. Anakwe, R.E., Jenkins, P.J. & Moran, M., 2011. Predicting Dissatisfaction After Total Hip Arthroplasty: A Study of 850 Patients. *The Journal of Arthroplasty*, 26(2), pp.209–213.
8. Anderson, A Shane, Loeser, R.F, 2010. Why is osteoarthritis an age-related disease? *Best Practice and Research: Clinical Rheumatology,* 24, pp.15–26
9. Arain, M. et al., 2010. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC medical research methodology*, 10(1), p.67.
10. Arden, N.K. et al., 2011. What is a good patient reported outcome after total hip replacement? *Osteoarthritis and Cartilage*, 19(2), pp.155–162.
11. ARUK & Public Health England. 2014. *Prevalence of osteoarthritis in England and local authorities: City of London*. Report.
12. Axford, J. et al., 2010. Prevalence of anxiety and depression in osteoarthritis: use of the Hospital Anxiety and Depression Scale as a screening tool. *Clinical Rheumatology*, 29(11), pp.1277–1283.
13. Ayers, D.C. et al., 2005. Total Knee Replacement Outcome and Coexisting Physical and Emotional Illness. *Clinical orthopaedics and related research*, 44; pp.157–161.
14. Azur, M.J. et al., 2011. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1), pp.40–49.

15. Bachmeier, C.J.M. et al., 2001. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. *Osteoarthritis and Cartilage*, 9(2), pp.137–146.
16. Bahadori, S. et al., 2018. A review of wearable motion tracking systems used in rehabilitation following hip and knee replacement. *Journal of Rehabilitation and Assistive Technologies Engineering*, 5; pp.1-8
17. Baker, P.N. et al., 2007. The role of pain and function in determining patient satisfaction after total knee replacement: DATA FROM THE NATIONAL JOINT REGISTRY FOR ENGLAND AND WALES. *Journal of Bone & Joint Surgery, British Volume*, 89-B(7), pp.893–900.
18. Bandura, A., 1977. *Social learning theory*, Prentice Hall ; Toronto : Prentice-Hall of Canada.
19. Barlow, T. et al., 2016. The effect of expectation on satisfaction in total knee replacements: a systematic review. *Springerplus*, 5(1), pp.167–167.
20. Barwais, F.A. et al., 2013. ActiGraph GT3X determined variations in “free-living” standing, lying, and sitting duration among sedentary adults. *Journal of Sport and Health Science*, 2(4), pp.249–256.
21. Bay, S. et al. 2018. A systematic review of psychological interventions in total hip and knee arthroplasty. *BMC Musculoskeletal Disorders* 19:201; pp. 1-11
22. Bedson, J. & Croft, P.R., 2008. The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature. *BMC Musculoskeletal Disorders*, 9(1), p.116.
23. Behrend, H. et al.; 2012. The “forgotten joint” as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplasty*. 27: pp.430–436. e431.
24. Bell, M.L. & Fairclough, D.L., 2014. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Statistical Methods in Medical Research*, 23(5), pp.440–459.
25. Bell, M.L. et al., 2014. Handling missing data in RCTs; a review of the top medical journals. *BMC medical research methodology*, 14; pp.118.
26. Berenbaum, F., 2013. Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). *Osteoarthritis and Cartilage*, 21(1), pp.16–21.
27. Berge, D.J. et al., 2004. Pre-operative and post-operative effect of a pain management programme prior to total hip replacement: a randomized controlled trial. *Pain*;110: pp. 33–39
28. Bishop, G.B. 1987. Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*. 51; pp.220-232
29. Blackburn, J. et al., 2012. Does preoperative anxiety and depression predict satisfaction after total knee replacement? *The Knee*, 19(5), pp.522–524.
30. Blackburn, S. et al., 2018. The extent, quality and impact of patient and public involvement in primary care research: a mixed methods study. *Research involvement and engagement*, 4(1), p.35.

31. Blagojevic, M. et al., 2010. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. *Osteoarthritis and Cartilage*, 18(1), pp.24–33.
32. Blom, A. et al. 2016. Improving patients' experience and outcome of total joint replacement: the RESTORE programme. *Programme grants for applied research.* Volume 4, Issue 12
33. Bouhlila, D.S. & Sellaouti, F., 2013. Multiple imputation using chained equations for missing data in TIMSS: a case study. *Large-scale Assessments in Education*, 1(1), pp.273–33.
34. Bourne, R.B., 2008. Measuring Tools for Functional Outcomes in Total Knee Arthroplasty. *Clinical Orthopaedics and Related Research*, 466(11), pp.2634–2638.
35. Bourne, R.B. et al., 2009. Patient Satisfaction after Total Knee Arthroplasty: Who is Satisfied and Who is Not? *Clinical Orthopaedics and Related research*, 468(1), pp.57–63.
36. Boynton, P.M. & Greenhalgh, T., 2004. Selecting, designing, and developing your questionnaire. *BMJ*, 328(7451), pp.1312–1315.
37. Brander, V. et al., 2007. The John Insall Award: Pain and Depression Influence Outcome 5 Years after Knee Replacement Surgery. *Clinical Orthopaedics and Related Research*. 464; pp. 21-26
38. Brander, V.A. et al., 2003. Ranawat Award Paper: Predicting Total Knee Replacement Pain. *Clinical Orthopaedics and Related Research*, 416, pp.27–36.
39. Brandes, M. et al., 2010. Changes in physical activity and health-related quality of life during the first year after total knee arthroplasty. *Arthritis Care & Research*, 63(3); pp. 328-334
40. Bryan, S. et al., 2018. Revisiting patient satisfaction following total knee arthroplasty: a longitudinal observational study. *BMC Musculoskeletal Disorders*, 19(1), pp. 1-8.
41. Bullens, P.H.J. et al., 2001. Patient satisfaction after total knee arthroplasty. *The Journal of Arthroplasty*, 16(6), pp.740–747.
42. Cai, L. et al., 2018. Incidence and Risk Factors of Kinesiophobia After Total Knee Arthroplasty in Zhengzhou, China: A Cross-Sectional Study. *The Journal of Arthroplasty*, 33(9), pp.2858–2862.
43. Carpenter, J.R., Kenward, M.G. & White, I.R., 2007. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Statistical Methods in Medical Research*, 16(3), pp.259–275.
44. Caspersen, C.J., Powell, K.E. & Christenson, G.M., 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports (Washington, D.C.);* 100(2), pp.126–131.
45. Chalmers, I., 1986. Minimizing Harm and Maximizing Benefit During Innovation in Health Care: Controlled or Uncontrolled Experimentation? *Birth*, 13(3), pp.155–164.
46. Chan, A.-W. & Altman, D.G., 2005. Epidemiology and reporting of randomised trials published in PubMed journals. *The Lancet*, 365(9465), pp.1159–1162.

47. Charlesworth, G. et al., 2013. Acceptance checklist for clinical effectiveness pilot trials: a systematic approach. *BMC medical research methodology*, 13(1), pp. 1-7.
48. Chen, K.Y. & Bassett, D.R Jr; 2005. The Technology of Accelerometry-Based Activity Monitors: Current and Future. *Medicine & Science in Sports & Exercise*, 37(Supplement), pp.S490–S500.
49. Choi, Y.-J. & Ra, H.J., 2016. Patient Satisfaction after Total Knee Arthroplasty. *Knee surgery & related research*, 28(1), pp.1–15.
50. Cirillo, D.J. et al., 2006. Effect of hormone therapy on risk of hip and knee joint replacement in the Women's Health Initiative. *Arthritis & Rheumatism*, 54(10), pp.3194–3204.
51. Cleeland, C.S., 1991. *Assessing Pain with the Brief Pain Inventory*,
52. Clement, N.D. et al. 2013. Primary total knee replacement in patients with mental disability improves their mental health and knee function: a prospective study. *Bone & Joint Journal*, 95-B(3), pp.360–366.
53. Clement, N.D. et al. 2013. Total knee replacement in patients with concomitant back pain results in a worse functional outcome and a lower rate of satisfaction. *Bone Joint Journal;* 95-B:pp. 1632–9.
54. Coggon, D. et al., 2000. Occupational physical activities and osteoarthritis of the knee. *Arthritis & Rheumatism*, 43(7), pp.1443–1449.
55. Cooper, C. et al., 1998. Individual risk factors for hip osteoarthritis: Obesity, hip injury and physical activity. *American Journal of Epidemiology*, 147(6), pp.516–522.
56. Cooper, C. et al., 1996. Occupational activity and the risk of hip osteoarthritis. *Annals of the Rheumatic Diseases*, 55(9), pp.680–682.
57. Cooper, C. et al., 2000. Risk factors for the incidence and progression of radiographic knee osteoarthritis. *Arthritis & Rheumatism*, 43(5), pp.995–1000.
58. Curran, D. et al., 1998. Incomplete quality of life data in randomized trials: missing forms. *Statistics in medicine*, 17(5-7), pp.697–709.
59. Dagenais, S. et al; 2008. Systematic Review of the Prevalence of Radiographic Primary Hip Osteoarthritis. *Clinical Orthopaedics and Related Research*, 467(3), pp.623–637.
60. Dawson, J. et al., 1998. Questionnaire on the perceptions of patients about total knee replacement. *Journal of Bone & Joint Surgery, British Volume*, 80(1), pp.63–69.
61. Dawson, J. et al. 1996. Questionnaire on the perceptions of patients about total hip replacement. *Journal of Bone & Joint Surgery, British Volume*, 78(2), pp.185–190.
62. de Groot, I.B. et al., 2008. Small increase of actual physical activity 6 months after total hip or knee arthroplasty. *Clinical Orthopaedics and Related Research*, 466(9), pp.2201–2208.
63. de Klerk, B.M. et al., 2009. No clear association between female hormonal aspects and osteoarthritis of the hand, hip and knee: a systematic review. *Rheumatology (Oxford, England)*, 48(9), pp.1160–1165.

64. de Lange-Brokaar, B.J.E. et al., 2012. Synovial inflammation, immune cells and their cytokines in osteoarthritis: a review. *Osteoarthritis and Cartilage*, 20(12), pp.1484–1499.
65. Department of Health. 2011, *Start active, stay active: A report on physical activity for health from the four home countries' Chief Medical Officers*
66. Dillon, C.F et al. 2006; Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991-94. *The Journal of Rheumatology;* 33 (11), pp. 2271-2279;
67. Dixon, K.E. et al.; 2007. Psychological Interventions for Arthritis Pain Management in Adults: A Meta-Analysis. *Health Psychology*, Vol. 26, No. 3, 241–250
68. Domecq, J.P. et al., 2014. Patient engagement in research: a systematic review. *BMC health services research*, 14(1), p.209.
69. Dorr, L.D. & Chao, L., 2007. The Emotional State of the Patient after Total Hip and Knee Arthroplasty. *Clinical Orthopaedics and Related Research*, 463; pp. 7-12.
70. Doury-Panchout, F. et al., 2015. Kinesiophobia negatively influences recovery of joint function following total knee arthroplasty. *European journal of physical and rehabilitation medicine*, 51(2), pp.155–161.
71. Dunbar, M.J. et al., 2013. I can't get no satisfaction after my total knee replacement: rhymes and reasons. *Bone & Joint Journal*, 95-B(Supplement), pp.148–152.
72. Edwards, H.B. et al., 2018. The Effect of Age, Sex, Area Deprivation, and Living Arrangements on Total Knee Replacement Outcomes. *JBJS Open Access*, 3(2), pp. 1-9
73. Edwards, M.L; Smith, B.C; 2011. The effects of the neutral response option on the extremeness of participant responses. *Incite Journal of Undergraduate Scholarship; Psychology, Research;* Volume 6
74. Eekhout, I. et al., 2014. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *Journal of Clinical Epidemiology*, 67(3), pp.335–342.
75. Escobar, A. et al., 2007. Effect of patient characteristics on reported outcomes after total knee replacement. *Rheumatology*, 46(1), pp.112–119.
76. EQ-5D-3L User Guide. EuroQol Group; 2013. http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/ EQ-5D-3L_UserGuide_2013_v5.0_October_2013.pdf [Accessed 17th March 2019]).
77. Evans, J.T. et al. 2019. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet*. 393; pp. 655–663
78. Eymir, M. et al., 2019 The effect of pain and kinesiophobia level on postoperative functional outcomes in patients with total knee arthroplasty. *BJJ Orthopaedic Proceedings.* Vol. 101-B, No. Supplement 5

79. Fairclough, D.L. et al., 1998. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Statistics in medicine*, 17(5-7), pp.781–796.
80. Fayers, P.M. et al.,1998. Incomplete quality of life data in randomized trials: missing items. *Statistics in medicine*, 17(5-7), pp.679–696.
81. Felson, D.T. et al., 1997. Risk factors for incident radiographic knee osteoarthritis in the elderly: the Framingham Study. *Arthritis & Rheumatism*, 40(4), pp.728–733.
82. Felson, D.T. et al., 1992. Weight loss reduces the risk for symptomatic knee osteoarthritis in women. The Framingham Study. *Annals of Internal Medicine*, 116(7), pp.535–539.
83. Filardo, G. et al., 2015. Patient kinesiophobia affects both recovery time and final outcome after total knee arthroplasty. *Knee Surgery, Sports Traumatology, Arthroscopy*, 24(10), pp.3322–3328.
84. Forsythe, M.E. et al., 2008. Prospective relation between catastrophizing and residual pain following knee arthroplasty: two-year follow-up. *Pain research & management : the journal of the Canadian Pain Society = journal de la société canadienne pour le traitement de la douleur*, 13(4), pp.335–341.
85. Franklin, P.D. et al., 2006. Pilot study of methods to document quantity and variation of independent patient exercise and activity after total knee arthroplasty. *The Journal of Arthroplasty*, 21(6 Suppl 2), pp.157–163.
86. Giesinger, K et al.; 2014. Comparative responsiveness of outcome measures for total knee arthroplasty. *Osteoarthritis Cartilage*; 22(2): pp184–189.
87. Gill, S.D. & McBurney, H., 2013. Does exercise reduce pain and improve physical function before hip or knee replacement surgery? A systematic review and meta-analysis of randomized controlled trials. *Archives of Physical Medicine and Rehabilitation*, 94(1), pp.164–176.
88. Gillon, R., 1994. Recruitment for clinical trials: the need for public-professional co-operation. *Journal of Medical Ethics*, 20(1), pp.3–4.
89. Golsorkhi, M. & Steel, R., 2018. *Report of the Patient Research Experience Survey 2017/18* 1st ed.
90. Grotle, M. et al., 2008. Obesity and osteoarthritis in knee, hip and/or hand: an epidemiological study in the general population with 10 years follow-up. *BMC Musculoskeletal Disorders*, 9(1), pp.132–132.
91. Güney-Deniz, H. et al., 2017. Does kinesiophobia affect the early functional outcomes following total knee arthroplasty? *Physiotherapy Theory and Practice*, 33(6), pp.448–453.
92. Gunston, F.H; 1971. Polycentric knee arthroplasty: prosthetic simulation of normal knee movement. *J Bone Joint Surg Br* 53B(2):pp. 272-277
93. Haanstra, T.M. & van den Berg, T., 2012. Systematic review: Do patient expectations influence treatment outcomes in total knee and total hip arthroplasty? *Health and Quality of Life Outcomes;* 10:152: pp. 1-14
94. Hanley, B. et al., 2004. *Involving the public in NHS public health, and social care research: briefing notes for researchers*,

95. Harding, P. et al., 2013. Do Activity Levels Increase After Total Hip and Knee Arthroplasty? *Clinical orthopaedics and related research*, 472(5), pp.1502–1511.
96. Hardy, S.E., Allore, H. & Studenski, S.A., Missing Data: A Special Challenge in Aging Research. *Journal of the American Geriatrics Society*, 57(4), pp.722–729.
97. Harris WH. 1969. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am*.;51(4):737–55.
98. Hartley, S.M. et al., 2008. Hope, self-efficacy, and functional recovery after knee and hip replacement surgery. *Rehabilitation Psychology*, 53(4), pp.521–529.
99. Hayes, D.A. et al., 2011. Knee arthroplasty: a cross-sectional study assessing energy expenditure and activity. *ANZ Journal of Surgery*, 81(5), pp.371–374.
100. Hendrick, P. et al., 2009. Does physical activity change predict functional recovery in low back pain? Protocol for a prospective cohort study. *BMC Musculoskeletal Disorders*, 10(1), p.136.
101. Heuts, P.H.T.G. et al., 2004. Pain-related fear and daily functioning in patients with osteoarthritis. *Pain*, 110(1-2), pp.228–235.
102. Hochberg, M.C. et al., 1995. The Association of Body-Weight, Body Fatness and Body-Fat Distribution with Osteoarthritis of the Knee - Data From the Baltimore Longitudinal-Study of Aging. *Journal of Rheumatology*, 22(3), pp.488–493.
103. Höll, S. et al., 2018. Clinical outcome and physical activity measured with StepWatch 3™ Activity Monitor after minimally invasive total hip arthroplasty. *Journal of orthopaedic surgery and research*, 13(1), p.148.
104. Hunter, D.J. & Felson, D.T., 2006. Clinical review-Osteoarthritis, *BMJ*; 332; pp.639–42
105. Ibrahim, J.G. et al., 2012. Missing data in clinical studies: issues and methods. *Journal of Clinical Oncology*, 30(26), pp.3297–3303.
106. Inoue, K. et al., 2000; Prevalence of hip osteoarthritis and acetabular dysplasia in French and Japanese adults. *Rheumatology*, 39, pp.745–748
107. Insall, J.N et al.; 1989. “Rationale of the knee society clinical rating system,” *Clinical Orthopaedics and Related Research*, no. 248, pp. 13–14.
108. Jacobson, A.F., 2016. Guided imagery for Total knee replacement: a randomized, placebo-controlled pilot study. *J Altern Complement Med*; 22: pp. 563–575
109. Jiang, L. et al., 2015. The objective evaluation of daily physical activity for knee osteoarthritis patients after total knee arthroplasty. A prospective clinical trial using accelerometry. *Osteoarthritis and Cartilage*, 23, A368.
110. John, D & Freedson, P., 2012. ActiGraph and Actical Physical Activity Monitors. *Medicine & Science in Sports & Exercise*, 44, pp. S86–S89.

111. Johns, R., 2005. One Size Doesn't Fit All: Selecting Response Scales For Attitude Items. *Journal of Elections, Public Opinion & Parties*, 15(2), pp.237–264.
112. Johnson, A.J. et al., 2013. Patient dissatisfaction with rehabilitation following primary total knee arthroplasty. *The journal of knee surgery*, 26(6), pp.417–421.
113. Juhakoski, R. et al., 2008. Risk factors for the development of hip osteoarthritis: a population-based prospective study. *Rheumatology*, 48(1), pp.83–87.
114. Kaushal, S., 2014. Missing data in clinical trials: Pitfalls and remedies. *International journal of applied & basic medical research*, 4(Suppl 1), pp.S6–7.
115. Kellgren, J. & Lawrence, J., 1963a. *Atlas of standard radiographic arthritis*, Oxford University Press.
116. Kim, C. et al., 2014. Prevalence of Radiographic and Symptomatic Hip Osteoarthritis in an Urban United States Community: The Framingham Osteoarthritis Study. *Arthritis & Rheumatology*, 66(11), pp.3013–3017.
117. Kirkley, A. et al., 2008. A Randomized Trial of Arthroscopic Surgery for Osteoarthritis of the Knee. *New England Journal of Medicine*, 359(11), pp.1097–1107.
118. Kitzinger, J., 1995. Qualitative research. Introducing focus groups. *BMJ: British Medical Journal*. 311; pp.299-302
119. Kitzinger, J., 1994. The methodology of Focus Groups: the importance of interaction between research participants. *Sociology of Health and Illness*, 16(1), pp.103–121.
120. Knight, S.R et al. 2011. Total Hip Arthroplasty - over 100 years of operative history. *Orthopedic Reviews*; 3: pp.72-74
121. Kori, S.H. et al.,1990. Kinisophobia: A new view of chronic pain behavior. *Pain Management*, 3(1), pp.35–43.
122. Krenk, L. et al., 2013. Activity, sleep and cognition after fast-track hip or knee arthroplasty. *The Journal of Arthroplasty*, 28(8), pp.1265–1269.
123. Krosnick, J.A. et al., The impact of "no opinion" response options on data quality: non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*.
124. Krych, A.J. et al., 2014. Does arthroscopic knee surgery work? *Arthroscopy : the Journal of Arthroscopic & related surgery;* 30(5), pp.544–545.
125. Kurtulus, D. et al., 2017. Kinesiophobia After Total Knee Arthroplasty. *Rheumatology*, 56, pp.162–162.
126. Lancaster, G.A., 2015. Pilot and feasibility studies come of age! *Pilot and Feasibility Studies*, 13, p.78.
127. Lancaster, G.A. et al., 2004. Design and analysis of pilot studies: recommendations for good practice. *Journal of evaluation in clinical practice*, 10(2), pp.307–312.
128. Lanyon, P. et al., 2000. Assessment of a genetic contribution to osteoarthritis of the hip: sibling study. *BMJ (British Medical Journal)*, 321(7270), pp.1179–1183.

129. Lareau, C. 2008. The Use of Accelerometers to Predict Functional Outcome After Total Knee Arthroplasty. *University of Massachusetts Medical School. Senior Scholars Program.* Paper 65.
130. Larsson, C. et al., 2014. Psychometric properties of the Tampa Scale of Kinesiophobia (TSK-11) among older people with chronic pain. *Physiotherapy Theory and Practice*, 30(6), pp.421–428.
131. Lavand'homme, P. & Thienpont, E., 2015. Pain after total knee arthroplasty: a narrative review focusing on the stratification of patients at risk for persistent pain. *The Bone & Joint Journal*, 97-B(10 Suppl A), pp.45–48.
132. Lawrence, R.C. et al., 2008. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis & Rheumatism*, 58(1), pp.26–35.
133. Learmonth, I.D. et al., 2007. The operation of the century: total hip replacement. *The Lancet*, 370(9597), pp.1508–1519.
134. Leite, W. & Beretvas, S.N., 2010. The Performance of Multiple Imputation for Likert-type Items with Missing Data. *Journal of Modern Applied Statistical Methods*, 9(1), pp.64–74.
135. Leon, A.C. et al., 2011. The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, 45(5), pp.626–629.
136. Leurent, B. et al., 2018. Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis: A Tutorial. *PharmacoEconomics*, 36(8), pp.889–901.
137. Lievense, A.M., 2002. Influence of obesity on the development of osteoarthritis of the hip: a systematic review. *Rheumatology*, 41(10), pp.1155–1162.
138. Lievense, A.M. et al., 2003. Influence of sporting activities on the development of osteoarthritis of the hip: A systematic review. *Arthritis & Rheumatism*, 49(2), pp.228–236.
139. Lingard, E.A. et al., 2004. Predicting the outcome of total knee arthroplasty. *The Journal of bone and joint surgery American volume*, 86-A(10), pp.2179–2186.
140. Lingard, E.A., Sledge, C.B. & Learmonth, I.D., 2006. Patient expectations regarding total knee arthroplasty: differences among the United States, United Kingdom, and Australia. *The Journal of Bone and Joint Surgery (American)*, 88(6), pp.1201–1207.
141. Little, R.J., D'agostino, R. & Cohen, M.L., 2012b. The prevention and treatment of missing data in clinical trials. *N Eng J Med.* 367(14); pp. 1355-1360
142. Little, R.J.A. & Rubin, D.B., 1987. Statistical analysis with missing data. *New York: Wiley*.
143. Litwic, A. et al., 2013. Epidemiology and burden of osteoarthritis. *British Medical Bulletin*, 105(1), pp.185–199.
144. Liu, S.-H. et al., 2016. Objectively Measured Physical Activity and Symptoms Change in Knee Osteoarthritis. *The American journal of medicine*, 129(5), pp.497–505.e1.
145. Lopes, V.P. et al., 2009. Actigraph calibration in obese/overweight and type 2 diabetes mellitus middle-aged to old adult patients. *Journal of physical activity & health*, 6 Suppl 1, pp.S133–40.

146. Loscalzo, J., 2009. Pilot trials in clinical research: of what value are they? *Circulation*, 119(13), pp.1694–1696.
147. M DiRusso, J., 2014. Variations in Delivery and Exercise Content of Physical Therapy Rehabilitation Following Total Knee Replacement Surgery: A Cross-Sectional Observation Study. *International Journal of Physical Medicine & Rehabilitation*, s5(01), pp.1–18.
148. Machin, D, Campbell, M.J., Tan, S.B, Tan, S.H. 2018. 'Sample Sizes for Clinical, Laboratory and Epidemiology Studies, Fourth Edition'. John Wiley and Sons.
149. MacGregor, A.J. & Spector, T.D., 1999. Twins and the genetic architecture of osteoarthritis. *Rheumatology*; 38: pp. 583–590
150. MacGregor, A.J. et al., 2000. The genetic contribution to radiographic hip osteoarthritis in women. *Arthritis & Rheumatism*; 43(11): pp. 2410–2416
151. MacGregor, A.J. et al., 2008. The genetic influence on radiographic osteoarthritis is site specific at the hand, hip and knee. *Rheumatology*, 48(3), pp.277–280.
152. Mahomed, N. et al., 2011. The Self-Administered Patient Satisfaction Scale for Primary Hip and Knee Arthroplasty. *Arthritis*; 4; pp.1–6.
153. Mancuso, C.A. et al., 2001. Patients' expectations of knee surgery. *The Journal of Bone and Joint Surgery (American)*, 83-A(7), pp.1005–1012.
154. Mancuso, C.A. et al., 1997. Patients' expectations and satisfaction with total hip arthroplasty. *Journal of Arthroplasty*, 12(4), pp.387–396.
155. Mancuso, C.A. et al., 2003. Patients with poor preoperative functional status have high expectations of total hip arthroplasty. *Journal of Arthroplasty*, 18(7), pp.872–878.
156. Mancuso, C.A. et al., 2008. Randomized trials to modify patients' preoperative expectations of hip and knee arthroplasties. *Clinical orthopaedics and related research*; 466: pp. 424–31.
157. Manninen, P. et al., 2002. Physical workload and the risk of severe knee osteoarthritis. *Scandinavian Journal of Work, Environment & Health*, 28(1), pp.25–32.
158. Mathie, M.J. et al., 2004. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement., 25(2), pp.R1–20.
159. Mathiessen, A. & Conaghan, P.G., 2017. Synovitis in osteoarthritis: current understanding with therapeutic implications. *Arthritis Research & Therapy*, 19(1), p.18.
160. McAlindon, T.E. et al., 1992. Radiographic patterns of osteoarthritis of the knee joint in the community: the importance of the patellofemoral joint. *Annals of the Rheumatic Diseases*, 51(7), pp.844–849.
161. Meding, J.B. et al., 2011. Pain Relief and Functional Improvement Remain 20 Years After Knee Arthroplasty. *Clinical orthopaedics and related research*, 470(1), pp.144–149.
162. Mercieca-Bebber, R. et al., 2016. Design, implementation and reporting strategies to reduce the instance and impact of missing

patient-reported outcome (PRO) data: a systematic review. *BMJ Open*, 6(6), p.e010938.
163. Miller, R.P. et al., 1991. The Tampa Scale: a Measure of Kinisophobia. *The Clinical journal of pain*, 7(1), p.51.
164. Mok, A. et al., 2019. Physical activity trajectories and mortality: population based cohort study. *BMJ*, 365, p.l2323.
165. Moran, M. et al., 2003. Expect the best, prepare for the worst: surgeon and patient expectation of the outcome of primary total hip and knee replacement. *Annals of The Royal College of Surgeons of England*, 85(3), pp.204–206.
166. Murray, D.W. et al., 2007. The use of the Oxford hip and knee scores. *Journal of Bone & Joint Surgery, British Volume*, 89(8), pp.1010–1014.
167. Nam, D. et al., 2014. Patient dissatisfaction following total knee replacement: a growing concern? *Bone & Joint Journal*, 96-B(Suppl), pp.96–100.
168. National Research Council, 2011. *The Prevention and Treatment of Missing Data in Clinical Trials*, National Academies Press.
169. Neogi, T. et al., 2009. Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. *BMJ*, 339; pp.b2844–b2844.
170. Nieuwenhuijsen, M.J., 2005. Design of exposure questionnaires for epidemiological studies. *Occupational and Environmental Medicine*, 62(4), pp.272–80– 212–4.
171. Nilsdotter, A.K.et al., 2009. Knee arthroplasty: are patients' expectations fulfilled? *Acta Orthopaedica*, 80(1), pp.55–61.
172. Osborne, J.W., 2012. *Best Practices in Data Cleaning*, SAGE.
173. Palazzo, C. et al., 2014. Determinants of satisfaction 1 year after total hip arthroplasty: the role of expectations fulfilment. *BMC Musculoskeletal Disorders*, 15(1), pp.53–53.
174. Palmer, K.T. et al., 2007. Does knee pain in the community behave like a regional pain syndrome? Prospective cohort study of incidence and persistence. *Annals of the Rheumatic Diseases*, 66(9), pp.1190–1194.
175. Pandit, H. et al., 2008. Pseudotumours associated with metal-on-metal hip resurfacings. *J Bone Joint Surg Br*. 90-B; pp.847-851.
176. Patel, J. et al. 2015. Predictors of Low Patient-Reported Outcomes Response Rates in the California Joint Replacement Registry. *Journal of Arthroplasty*, 30(12), pp.2071–2075.
177. Parvizi, J. et al., 2010. Back pain and total hip arthroplasty: a prospective natural history study. *Clinical Orthopaedics and Related Research*, 468(5), pp.1325–1330.
178. Paxton, R.J., 2015. Physical activity after total knee arthroplasty: A critical review. *World journal of orthopedics*, 6(8), pp.614–10.
179. Pedersen, A.B. et al., 2017. Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology*, 9, pp.157–166.
180. Peyre, H. et al., 2011. Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple

imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey. *Quality of Life Research*, 20(2), pp.287–300.
181. Picavet, H.S.J., 2002. Pain Catastrophizing and Kinesiophobia: Predictors of Chronic Low Back Pain. *American Journal of Epidemiology*, 156(11), pp.1028–1034.
182. Pigott, T.D., 2010. A Review of Methods for Missing Data. *Educational Research and Evaluation*, 7(4), pp.353–383.
183. Poquet, N. & Lin, C., 2016. The Brief Pain Inventory (BPI). *Journal of physiotherapy*, 62(1), p.52.
184. Porta, M. et al., 2014. *A Dictionary of Epidemiology*, Oxford University Press.
185. Pozzi, F. et al., 2013. Physical exercise after knee arthroplasty: a systematic review of controlled trials. *European journal of physical and rehabilitation medicine*, 49(6), pp.877–892.
186. Punzi, L. et al., 2016. Post-traumatic arthritis: overview on pathogenic mechanisms and role of inflammation. *RMD Open*; 2:e000279; pp. 1-9
187. Reginster, J.Y., 2002. The prevalence and burden of arthritis. *Rheumatology*, 41(90001), pp.3–6.
188. Reijman, M. et al., 2007. Body mass index associated with onset and progression of osteoarthritis of the knee but not of the hip: the Rotterdam Study. *Annals of the Rheumatic Diseases*, 66(2), pp.158–162.
189. Riddle, D.L. et al., 2011. Pain Coping Skills Training for Patients With Elevated Pain Catastrophizing Who Are Scheduled for Knee Arthroplasty: A Quasi-Experimental Study. *Archives of Physical Medicine and Rehabilitation*, 92(6), pp.859–865.
190. Robertsson, O. et al., 2000. Patient satisfaction after knee arthroplasty: a report on 27,372 knees operated on between 1981 and 1995 in Sweden. 71(3), pp.262–267.
191. Rolstad, S. et al., 2011. Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 14(8), pp.1101–1108.
192. Rombach, I. et al., 2018. Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. *BMC medical research methodology*, 18(1), p.87.
193. Rombach, I. et al., 2016. The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*, 25(7), pp.1613–1623.
194. Roos, E. and Toksvig-Larsen, S., 2003. Knee injury and Osteoarthritis Outcome Score (KOOS) – validation and comparison to the WOMAC in total knee replacement. *Health and Quality of Life Outcomes*,1:17; pp. 1-10

195. Royston, P. & White, I.R., 2011a. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. *Journal of Statistical Software*, 45(4), pp.1–20.
196. Rubin, D.B., 2009. *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons.
197. Sacristan, J.A. et al., 2016. Patient involvement in clinical research: why, when, and how. *Patient Preference and Adherence*, pp.631–10.
198. Schafer, J.L., 1997. *Analysis of Incomplete Multivariate Data*, CRC Press.
199. Scott, C. et al., 2016. Predicting dissatisfaction following total knee arthroplasty in patients under 55 years of age. *J Bone Joint Surg [Br];* 2010;92-B: pp.1253-1258.
200. Scott, C.E.H. et al., 2012. Patient expectations of arthroplasty of the hip and knee. *The Journal of bone and joint surgery British volume*, 94(7), pp.974–981.
201. Seavey, W.G. et al., 2003. Risk factors for incident self-reported arthritis in a 20 year follow-up of the Alameda County Study Cohort. *The Journal of Rheumatology*. 30(10); pp. 2103-2111;
202. Simons, C.L. et al., 2015. Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? *Quality of Life Research*, 24(4), pp.805–815.
203. Skender, S. et al., 2016. Accelerometry and physical activity questionnaires - a systematic review. *BMC Public Health*, 16(1), p.126.
204. Skinner, J. et al., 2003. Racial, ethnic, and geographic disparities in rates of knee arthroplasty among Medicare patients. *New England Journal of Medicine*, 349(14), pp.1350–1359.
205. Smith-Petersen, M.N.; 1948. The classic: evolution of mould arthroplasty of the hip joint. *J Bone Joint Surg* 30B: L59.
206. Snyder, C.F. et al., 2007. Patient-reported outcome instrument selection: designing a measurement strategy. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 10 Suppl 2, pp.S76–85.
207. Söderman, P. et al.; 2001. Outcome of total hip replacement: a comparison of different measurement methods. *Clinical orthopaedics and related research* (390):163-72.
208. Somers, T.J. et al., 2009. Pain Catastrophizing and Pain-Related Fear in Osteoarthritis Patients: Relationships to Pain and Disability. *Journal of Pain and Symptom Management*, 37(5), pp.863–872.
209. Song, E-K; et al. 2013. Arthroplasty - Update, Plamen Kinov, IntechOpen, DOI: 10.5772/54343.
210. Spector, T.D. et al., 1994. Incidence and progression of osteoarthritis in women with unilateral knee disease in the general population: the effect of obesity. *Annals of the Rheumatic Diseases*, 53(9), pp.565–568.
211. Srikanth, V.K. et al., 2005. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis and Cartilage*, 13(9), pp.769–781.
212. Staley, K. et al., 2012. *INVOLVE: Developing training and support for public involvement in research*,

213. Stern, A.F., 2014. The hospital anxiety and depression scale. *Occupational medicine (Oxford, England)*, 64(5), pp.393–394.
214. Sterne, J.A.C. et al., 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338; pp.b2393–b2393.
215. Sullivan, M. et al., 2009. Psychological determinants of problematic outcomes following Total Knee Arthroplasty. *Pain*, 143(1), pp.123–129.
216. Sullivan, M.J.L. et al., 1995. Pain Catastrophizing Scale. User Manual. *PsycTESTS Dataset*.
217. Sutton, A.J. et al., 2001. A case-control study to investigate the relation between low and moderate levels of physical activity and osteoarthritis of the knee using data collected as part of the Allied Dunbar National Fitness Survey. *Annals of the Rheumatic Diseases*, 60(8), pp.756–764.
218. Tepper, S. & Hochberg, M.C., 1993. Factors Associated with Hip Osteoarthritis - Data From the 1st National-Health and Nutrition Examination Survey (Nhanes-I). *American Journal of Epidemiology*, 137(10), pp.1081–1088.
219. Thewlis, D. et al., Objectively measured 24-hour activity profiles before and after total hip arthroplasty. *Bone & Joint Journal*. 101-B: pp.415–425.
220. Thijssen, E., van Caam, A. & van der Kraan, P.M., 2014. Obesity and osteoarthritis, more than just wear and tear: pivotal roles for inflamed adipose tissue and dyslipidaemia in obesity-induced osteoarthritis. *Rheumatology (Oxford, England)*, 54(4), pp.588–600.
221. Thornton, H., 2008. Patient and public involvement in clinical trials. *BMJ*, 336(7650), pp.903–904.
222. Tkachuk, G.A. & Harris, C.A., 2012. Psychometric properties of the Tampa Scale for Kinesiophobia-11 (TSK-11). *The journal of pain: official journal of the American Pain Society*, 13(10), pp.970–977.
223. Tolk, J.J. et al., 2018. The EKSPECT study: the influence of Expectation modification in Knee arthroplasty on Satisfaction of PatiEnts: study protocol for a randomized Controlled Trial. *Trials* volume 19, Article number: 437; pp. 1-10
224. Toogood, P.A. et al., 2016. The monitoring of activity at home after total hip arthroplasty. *The bone & joint journal*, 98-B(11), pp.1450–1454.
225. Tortora, G.J. & Derrickson, B.H., 2010. *Principles of Anatomy and Physiology, 13th Edition*, Wiley Global Education.
226. Troiano, R.P. et al., 2008. Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise*, 40(1), pp.181–188.
227. Tsonga, T. et al., 2011. Evaluation of improvement in quality of life and physical activity after total knee arthroplasty in greek elderly women. *The open orthopaedics journal*, 5(1), pp.343–347.
228. van Buuren, S., 2007b. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*.

229. van den Akker-Scheek, I. et al., 2007. Preoperative or postoperative self-efficacy: Which is a better predictor of outcome after total hip or knee arthroplasty? *Patient Education and Counseling*, 66(1), pp.92–99.
230. Vasarhelyi, E.M. & MacDonald, S.J., 2012. The influence of obesity on total joint arthroplasty. *Journal of Bone & Joint Surgery, British Volume*, 94-B(11_Supple_A), pp.100–102.
231. Vissers, M.M. et al., 2010. Functional capacity and actual daily activity do not contribute to patient satisfaction after total knee arthroplasty. *BMC Musculoskeletal Disorders*, 11(1), p.108.
232. Vissers, M.M. et al., 2013. Physical functioning four years after total hip and knee arthroplasty. *Gait & posture*, 38(2), pp.310–315.
233. Waldrop, D. et al., 2001. Self-efficacy, optimism, health competence, and recovery from orthopedic surgery. *Journal of Counseling Psychology*, 48(2), pp.233–238.
234. Walker, D.J. et al., 2002. Measured ambulation and self-reported health status following total joint replacement for the osteoarthritic knee. *Rheumatology*, 41(7), pp.755–758.
235. Wallis, J.A. et al., 2013. What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis. *Osteoarthritis and Cartilage*, 21(11), pp.1648–1659.
236. Westby, M.D., 2012. Rehabilitation and Total Joint Arthroplasty. *Clinics in geriatric medicine*, 28(3), pp.489–508.
237. White, I.R., Royston, P. & Wood, A.M., 2010. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in medicine*, 30(4), pp.377–399.
238. Whitehouse, S.L. et al., 2005. The Oxford Knee Score; problems and pitfalls. *The Knee*, 12(4), pp.287–291.
239. Wilder, F.V. et al., 2002. History of acute knee injury and osteoarthritis of the knee: a prospective epidemiological assessment. *Osteoarthritis and Cartilage*, 10(8), pp.611–616.
240. Williams, D.P. et al., 2013. Early postoperative predictors of satisfaction following total knee arthroplasty. pp.1–5.
241. Wisniewski, S.R. et al., 2006. Prevention of Missing Data in Clinical Research Studies. *Biological Psychiatry*, 59(11), pp.997–1000.
242. Witvrouw, E. et al., 2009. Catastrophic thinking about pain as a predictor of length of hospital stay after total knee arthroplasty: a prospective study. *Knee Surgery, Sports Traumatology, Arthroscopy*, 17(10), pp.1189–1194.
243. Wluka, A.E., 2009. Ethnicity may be a risk modifier for knee osteoarthritis. *International Journal of Clinical Rheumatology*, 4(1), pp.29–32.
244. Woby, S.R. et al., 2005. Psychometric properties of the TSK-11: a shortened version of the Tampa Scale for Kinesiophobia. *Pain*, 117(1-2), pp.137–144.
245. Wong, L.P., 2008. Focus group discussion: a tool for health and medical research. *Singapore medical journal*, 49(3), pp.256–60–quiz 261.

246. Wood, A.M. et al., 2004. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical trials (London, England)*, 1(4), pp.368–376.
247. Wylde, V. et al., 2007. Total knee replacement: Is it really an effective procedure for all? *The Knee*, 14(6), pp.417–423.
248. Yang, C.-C. & Hsu, Y.-L., 2010. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors*, 10(8), pp.7772–7788.
249. Zigmond, A.S. & Snaith, R.P., 1983. The hospital anxiety and depression scale. *Acta psychiatrica Scandinavica*, 67(6), pp.361–370.