Analysis of knee replacements using data from the National Joint Registry for England and Wales

Number of Volumes: 1

Paul Nicholas Baker

Doctor of Medicine

Institute of Cellular Medicine

Newcastle University

Date of Submission: June 2013

Abstract

Introduction: Establishing best practice for knee replacement is important given the large number of procedures performed. Research into knee replacement is problematic given that implant failure is a rare event. The logistical and financial costs associated with prospective clinical trials are therefore high. Research using national arthroplasty registers may overcome some of these difficulties.

Aim: To assess whether research performed on data recorded by the National Joint Registry for England and Wales has the ability to answer clinically relevant research questions relating to knee replacement surgery. To determine if registry research is able to answer specific clinical questions that are unsuited to prospective randomised clinical trial designs.

Methods: Analyses was performed using combined data from the National Joint Registry for England and Wales (NJR) and the Department of Health Patient Reported Outcome Measures (PROMs) project.

Results: Nine specific analyses investigated the ability of registry data to ask pertinent clinical questions relating to three areas of practice: unicondylar knee replacement (UKR), total knee replacement (TKR), revision knee replacement (RTKR).

Discussion: Registry analyses are well suited to the analysis of rare outcomes such as implant revision and death. In comparison to prospective clinical trial designs they are cheaper, consume less time and resources and have the ability to identify associations and additional factors that may potentially influence outcome. As they use current national data they are more representative of "real-time" national practice and as such overcome some of the problems of generalisability associated with more rigidly designed clinical trials. However, as no information is collected about clinical decision making, drawing strong causal inferences from this type of data is problematic.

Conclusion: Using registry data it is possible to answer a range of clinically important research questions. However, due to their limitations, it is necessary to combine information from these observational databases with clinical trial data before robust recommendations that influence clinical practice can be made. The key question researchers have to answer now is how registry data and clinical trial data can be effectively integrated.

Dedication

I dedicate this work to my wife Rosie and my children George, Henry and Percy. Without their love, support and understanding I would not have been able to undertake and complete this body of work.

Acknowledgements

Thanks to Professor David Deehan and Professor Paul Gregg. Their support and guidance has been invaluable throughout the MD process.

I would also like to acknowledge the input and help received from Simon Jameson and Mike Reed who aided in the development of a number of the research ideas presented. I am also thankful to Peter Avery and Professor Steve Rushton from Newcastle University for their statistical guidance.

I would like to thank the patients and staff of all of the hospitals in England and Wales who have contributed data to the National Joint Registry. I am grateful to the Healthcare Quality Improvement Partnership, the National Joint Registry Steering Committee, and the staff at the National Joint Registry Centre for facilitating this work. This work was funded by a fellowship from the National Joint Registry. Throughout the process I have conformed to the National Joint Registry's standard protocol for data access and publication.

The views expressed in this thesis do not necessarily reflect the views of the National Joint Registry Steering Committee or the Health Quality Improvement Partnership (HQIP) who do not vouch for the results and conclusions derived from any of the analyses performed as part of this doctoral thesis.

Contents

Declaration

I confirm this is an original piece of work that has not been submitted for consideration of a higher degree award with another institution or university. This work has been carried out by the candidate under the supervision of the supervising research team as per the guidance for doctoral thesis laid out by Newcastle University. The role and input of each member of the supervisory team and any associated research collaborators is detailed in the statement of attribution.

Statement of Attribution

The role and input of each person contributing to this research is listed below:

Prof David Deehan:

Prof Deehan is the Professor of orthopaedic surgery at Newcastle University. He was the lead university supervisor for this thesis who had responsibility for my day to day supervision. He introduced me to researchers within Newcastle University who were able to assist me during my thesis. He helped identify clinically relevant research questions (Analyses 8, 9) and assisted with the interpretation of data generated during analysis of the registry data. He also helped with the drafting and revision of the thesis and all of the associated publications that derived from this thesis.

Prof Paul Gregg:

Prof Gregg is the Professor of orthopaedic surgery at the James Cook University Hospital, Middlesbrough and the current vice-chairman of the National Joint Registry for England and Wales (NJR). He was my supervisor within the NJR. In this capacity he helped resolve any administrative difficulties that arose within the NJR. These included issues such as data access and support from the NJR data handler. He helped identify clinically relevant

research questions (Analyses 1, 3, and 4) and assisted with the interpretation of data generated during analyses of the registry data relating to these questions. He also helped with the drafting and revision of the publications derived from this thesis.

Mr Mike Reed:

Mr Mike Reed is a consultant orthopaedic surgeon at Wansbeck General Hospital, Ashington. He helped identify clinically relevant research questions (Analysis 5) and assisted with the interpretation of data generated during analyses of the registry data relating to these questions.

Simon Jameson:

Simon Jameson worked with me as the NJR research fellow between April 2011 and April 2012. He provided statistical and methodological support and review for work presented in this thesis.

Prof Stephen Rushton:

Prof Rushton is a Professor of biological modelling at Newcastle University. Prof Ruhston was involved in the direct analysis of data presented in analysis 6. He undertook the ordinal regression and structured equational modelling (SEM) presented as part of this analysis and helped interpret the results of these analyses. He helped write the methodological section 3.5.4 outlining the SEM approach to analysis and provided figure 3.11.

Dr Peter Avery:

Dr Avery is a lecturer in mathematics and statistics at Newcastle University. He helped with the interpretation of statistical output for analyses 1, 2 and 5.

All other work relating to this thesis was undertaken by me. This includes the generation and development of clinically relevant research questions for analyses 2, 6, 7. I took sole responsibility for the receipt, storage, cleaning and handling of all NJR data. Other than Prof Stephen Rushton (analysis 6) no other party had access to this data. I designed and subsequently undertook all of the statistical analyses presented in the thesis with the exception of analysis 6. This thesis was written by me without the assistance of any third party other than the previously described contributions of Prof Rushton.

Ethics Statement

Prospective projects were discussed with the chairs of the local ethics and research and development committees. Correspondence with Paddy Stephenson chair of the Sunderland local research and ethic committee can be found the appendices. As no patient identifiable data was requested and as no additional patient contact was required it was decided that all projects could be performed as service evaluations without the need for additional ethical approval. The NJR has its own consent mechanism for data collection and it was felt that this was sufficient for the purposes of the proposed research.

List of Tables and Figures

Tables

Figures

Papers and Presentations derived from this thesis

Papers

Baker P, Petheram T, Jameson S, Avery P, Reed M, Gregg P, Deehan D. Comparison of patient-reported outcome measures following total and unicondylar knee replacement. *Journal of Bone and Joint Surgery (Br)* 2012;94(7):919-27.

Baker P, Petheram T, Avery P, Gregg P, Deehan D. Revision for unexplained pain following Unicompartmental and Total Knee Replacement. *Journal of Bone and Joint Surgery (Am)* 2012;94(17);e126 1-7.

Baker P, Jameson S, Deehan D, Gregg P, Porter M, Tucker K. Equivalent Survival of Medial and Lateral Unicondylar Knee Replacement: An analysis of data from a National Joint Registry. *Journal of Bone and Joint Surgery (Br)* 2012;94(12):1641-8.

Baker P, Jameson S, Critchley R, Reed M, Gregg P, Deehan D. Centre and surgeon volume influence the revision rate following unicondylar knee replacement: an analysis of 23,400 medial cemented unicondylar knee replacements. *Journal of Bone and Joint Surgery (Am)* 2013;95(8):702-9.

Baker P, Deehan D, Lees D, Jameson S, Avery P, Gregg P, Reed M. The effect of surgical factors on early Patient Reported Outcome Measures (PROMs) following total knee replacement. *Journal of Bone and Joint Surgery (Br)* 2012;94(8):1058-66.

Baker P, Rushton S, Jameson S, Reed M, Gregg P, Deehan D. Patient satisfaction with total knee replacement cannot be predicted from pre-operative variables alone: A cohort study from the National Joint Registry for England and Wales. *Bone and Joint Journal* 2013;95-B(10):1359-65.

Baker P, Petheram T, Jameson S, Reed M, Gregg P, Deehan D. The Association Between Body Mass Index and the Outcomes of Total Knee Arthroplasty. *Journal of Bone and Joint Surgery (Am)* 2012;94(16):1501-8.

Baker P, Cowling P, Kurtz S, Jameson S, Gregg P, Deehan D. Reason for Revision Influences Early Patient Outcomes After Aseptic Knee Revision. *Clinical Orthopaedics and Related Research* 2012;470(8);2244-52.

Baker P, Deehan D. CORR insights: Reason for revision TKA predicts clinical outcome: Prospective evaluation of 150 consecutive patients with 2-years follow-up. *Clinical Orthopaedics and Related Research* 2013;471(7):2303-4.

Baker P, Petheram TG, Kurtz S, Konttinen YT, Gregg P, Deehan D. Patient reported outcome measures after revision of the infected TKR: comparison of single *versus* two-stage revision. *Knee Surg Sports Traumatol Arthrosc*. 2013;21(12):2713-20.

Baker P, Critchley R, Jameson S, Gray A, Port A, Reed M, Deehan D. Revision knee replacement in England and Wales: An audit of hospital volume. *Bulletin of the Royal College of Surgeons of England* 2013;95(5):1-5.

Presentations

Comparison of Patients Reported Outcomes Measures following total and unicondylar knee replacement.

- British Society for Surgery of the Knee: Derby 2012 (Podium)
- EFORT: Berlin 2012 (Podium)
- International Society of Arthroplasty Registers: Bergen 2012 (Podium)

Revision for unexplained pain following Unicompartmental and Total Knee Replacement.

- American Academy of Orthopaedic Surgeons: San Francisco 2012 (Poster)
- British Society for Surgery of the Knee: Derby 2012 (Podium)

Equivalent Survival of Medial and Lateral Unicondylar Knee Replacement: An analysis of data from a National Joint Registry.

- International Society of Arthroplasty Registers: Bergen 2012 (Podium)
- British Orthopaedic Association: Manchester 2012 (Podium)

- American Academy of Orthopaedic Surgeons: Chicago 2013 (Podium)

Centre and surgeon volume influence the revision rate following unicondylar knee replacement: an analysis of 23,400 medial cemented unicondylar knee replacements.

- American Academy of Orthopaedic Surgeons: Chicago 2013 (Poster)

The effect of surgical factors on early Patient Reported Outcome Measures (PROMs) following total knee replacement.

- British Society for Surgery of the Knee: Derby 2012 (Podium)
- International Society of Arthroplasty Registers: Bergen 2012 (Podium)

The influence of patient factors on patient's perceptions of satisfaction and success following total knee replacement.

- British Orthopaedic Association: Manchester 2012 (Podium)
- American Academy of Orthopaedic Surgeons: Chicago 2013 (Poster)

Is it justified to ration total knee arthroplasty based on Body Mass Index?

- British Society for Surgery of the Knee: Derby 2012 (Podium)
- EFORT: Berlin 2012 (Poster)

Reason for Revision Influences Early Patient Outcomes after Aseptic Knee Revision.

- American Academy of Orthopaedic Surgeons: San Francisco 2012 (Podium)

Revision knee replacement in England and Wales: An audit of hospital volume.

- British Orthopaedic Association: Manchester 2012 (Podium)

Chapter 1: Introduction

1.1 Knee replacement surgery in 2013

1.1.1 The rising incidence of knee replacements

Over the last 50 years knee replacement has become the 'gold standard' for the treatment of end stage osteoarthritis of the knee. It is an effective way of relieving pain, restoring physical function and improving health related quality of life in this group of patients (Drewett 1992, Rissanen 1995, Jones 2000, Bachmeier 2001, Fitzgerald 2004). Changing social demographics and an increasingly aged population are causing a steady rise in the number of patients in need of knee replacement (Kurtz 2007, Lawrence 2008). As a result there has been a year on year increase in the number of knee replacements being performed in England and Wales over the last twenty years (Culliford 2010, NJR-AR 2012). This trend is mirrored in other developed countries including the United States (Kurtz 2007, Culliford 2010); Sweden, Denmark and Norway (Robertsson 2010); and South Korea (Kim 2008).

At the turn of the millennium it was estimated that approximately 20 million Americans had symptomatic osteoarthritis (Lawrence 1998) with more than 10 million suffering from osteoarthritis of the knee (Parmet 2003). The Framlington Osteoarthritis study (1983-1985) estimated the prevalence of radiographic knee osteoarthritis in patients aged 63-93 was approximately 33% (Lawrence 1998). Currently over 650,000 total knee replacements are performed annually in the United States (Carr 2012) and over 80,000 in England and Wales (NJR-AR 2012). These figures are expected to increase further over the next 20 years, representing an increasing public health problem, as the population of the world ages and medical interventions continue to increase average life expectancy (Birrel 1999, Ehrlich 2003, Kurtz 2007).

1.1.2 Knee joint failure: Why are knee replacements required?

In England and Wales, as in other developing countries, the commonest indication for primary knee replacement is end stage osteoarthritis (NJR-AR 2012). Osteoarthritis was the indication for surgery in approximately 95% of all knee replacements performed in England and Wales in 2011 (NJR-AR 2012). While the rates of knee replacement performed for osteoarthritis have increased over the last 10 years the number of knee replacements performed for rheumatoid and other inflammatory arthropathies has steadily declined. This is most likely due to the increasing use and efficacy of a range of biologically active disease modifying anti-inflammatory drugs (DMARDS) to treat these conditions (Louie 2010).

Osteoarthritis is a degenerative process that primarily affects articular cartilage and the underlying subchondral bone. It is the result of a combination of genetic, metabolic, biochemical and biomechanical factors with secondary components of inflammation (Iannone 2003). Its characteristic features include cartilage loss, peri-articular bone response with subchondral sclerosis and cysts, and the formation of osteophytes (Figure 1.1). Cartilage degradation in combination with a disordered repair process cause progressive joint damage. In the majority of patients, the initiating mechanism is damage to normal articular cartilage by physical forces, following either a single event of macrotrauma or repeated microtrauma (Mow 1992).

Knee osteoarthritis is the most prevalent form of osteoarthritis, one of the most common diseases affecting humans and a common cause of disability (Symmons 2002). A number of constitutional and mechanical factors influence the development of knee osteoarthritis (Carr 2012) the most important of which are age, obesity, previous trauma and gender (Oliveria 1995, Symmons 2002, Fehring 2007). These factors are important as they influence bone morphology, bone density, hormone levels, mechanical loads and the propensity to trauma all of which are implicated in the development of arthritis (Carr 2012). The incidence

of knee osteoarthritis is higher in woman compared to men and this difference is most marked with increasing age (Dequeker 1998, Oliveria 1995). Osteoarthritis and its treatment has significant cost and time implications for the NHS, accounting for up to 25% of visits to primary care providers, and half of all NSAIDs (Non-Steroidal Anti-Inflammatory Drugs) prescriptions (Green 2001).

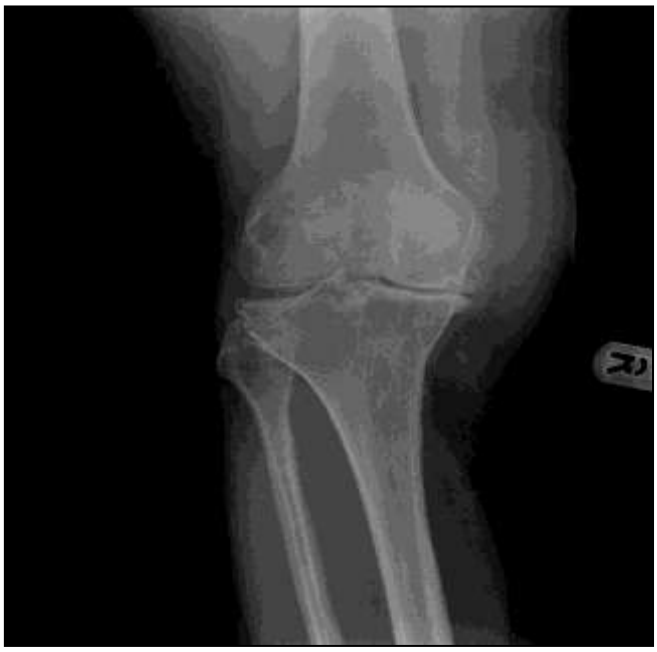


Figure 1.1: Radiographic representation of a knee with end-stage osteoarthritis.

1.1.3 When is surgery for knee arthritis indicated?

In 2007 the National Institute for Clinical Excellence (NICE) published guidance on the care and management of knee osteoarthritis as part of wider recommendations for the treatment of osteoarthritis in adults (NICE (CG59) 2007). This included specific guidance on the role of knee replacement surgery:

- Clinicians with responsibility for referring a person with knee osteoarthritis for consideration of surgery should ensure that the person has been offered at least the core (non-surgical) treatment options.

- Referral for knee replacement surgery should be considered for people with osteoarthritis who experience knee symptoms (pain, stiffness and reduced function) that have a substantial impact on their quality of life and are refractory to non-surgical treatment. Referral should be made before there is prolonged and established functional limitation and severe pain.
- Patient-specific factors (including age, gender, smoking, obesity and comorbidities) should not be barriers to referral for knee replacement surgery.

The role of knee replacement in rheumatoid/inflammatory arthritis is similar. The American College of Rheumatology currently recommend that 'In patients who have unacceptable levels of pain, loss of range of motion, or limitation of function because of structural joint damage, surgical procedures should be considered.' (American College of Rheumatology 2002). The consistent message from these guidelines is that knee replacement should be reserved for disease which has failed conservative treatment and which has a significant impact on the patient function, well-being and quality of life. In these patients with 'end-stage' disease knee replacement is currently the treatment of choice. In this group of patients it is an effective way of reducing pain, restoring functional capacity, and improving quality of life (Drewett 1992, Rissanen 1995, Fitzgerald 2004). These improvements occur irrespective of patient age (March 1999). It has been shown that approximately 32% of patients undergoing knee replacement return to a "normal" quality of life (a Quality of Life index of 1.00) following surgery (Norman-Taylor 1996).

1.1.4 Which surgical options are available for the treatment of knee osteoarthritis?

Surgeons have a number of different surgical options available for the treatment of patients with 'end stage' symptomatic knee osteoarthritis. Operative interventions can be broadly divided into those that preserve the knee joint (osteotomy and arthroscopic procedures) and those that replace the joint (arthroplasty/replacement procedures). The role of joint preserving procedures such as osteotomy and arthroscopy is limited in established

osteoarthritic disease. Osteotomy involves dividing the bone of the proximal tibia and realigning the weight bearing axis of the lower limb (Figure 1.2). This helps to 'off load' the affected portion of the joint and redistributes the applied loads through the preserved portion of the knee. While results suggest physical functioning after osteotomy is marginally better than the function observed after replacement procedures the higher rates of complication and inferior survivorship have meant this procedure is largely reserved for patients with significant mechanical malalignment due to peri-articular deformity (Stukenborg-Colsman 2001, Brouwer 2007).

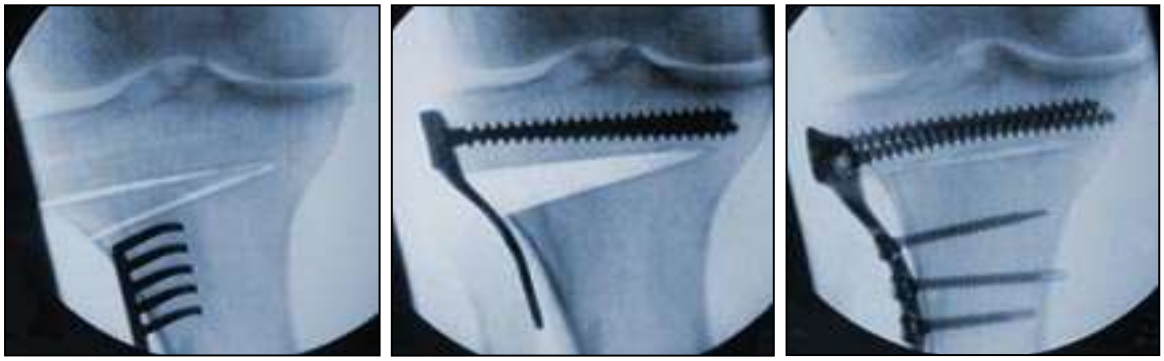


Figure 1.2: Intra-operative sequence demonstrating a high tibial osteotomy to re-align the mechanical axis of the knee joint.

Recent NICE guidelines on the role of knee arthroscopy state that arthroscopic knee washout alone should not be used as a treatment for osteoarthritis because it does not demonstrate clinically useful benefit in the short or long term (NICE (IPG 230) 2007). NICE recommends its use is limited to the small number of patients who have knee osteoarthritis and a clear history of mechanical locking, not including those knees with isolated gelling, 'giving way' or loose bodies (NICE (IPG 230) 2007).

The majority of patients who require surgery having exhausted non-operative measures will therefore undergo knee joint replacement. Once the decision is made to replace the knee

the surgeon has to determine whether the patient should undergo partial (replacement of only one section of the knee using a unicondylar (UKR) or patello-femoral (PFR) replacement) or total (replacement of the whole tibio-femoral joint with or without patella resurfacing) knee replacement (TKR) and which brand of replacement should be used (Figure 1.3).

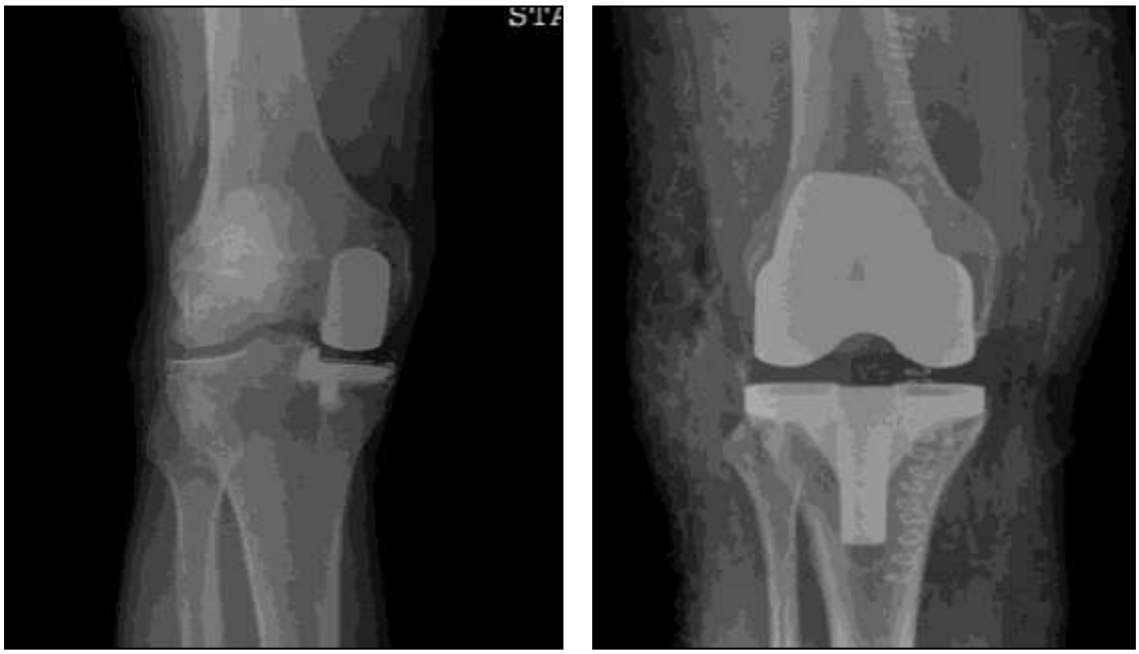
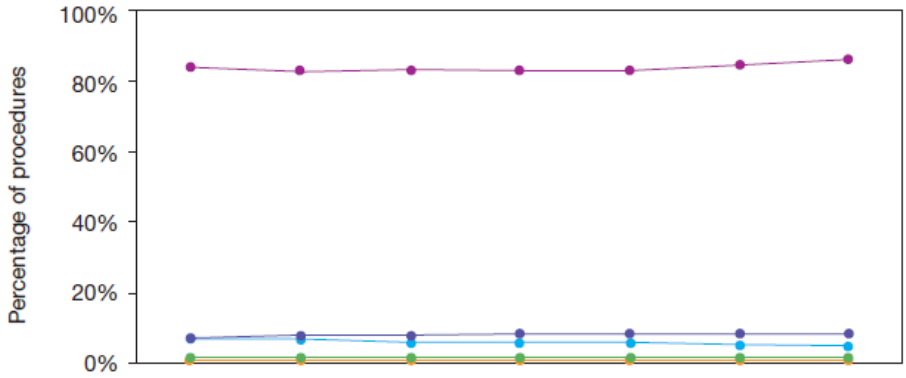


Figure 1.3: Radiographs of a Unicondylar (left) and Total (right) knee replacement.

The indications for both partial and total knee replacement are similar. However, while total knee replacement replaces the whole joint partial knee replacement only replaces the affected compartments of the knee leaving the rest of the knee intact. Perceived advantages of partial replacements are that they are a less invasive procedure; produce better functional outcomes and are associated with easier revision when they eventually fail. While there is some supportive data from single surgeons/centres with surgical enthusiasm for this approach (Laurecin 1991, Newman 2009, Pandit 2011) the benefits of partial replacement at a national level remain to be proven in large scale clinical trials. In addition partial replacement is consistently associated with a significantly higher rate of failure when compared to total knee replacement (AJR-AR 2009, Labek 2011, NJR-AR 2012).

Encouraging reports from a number of small series have seen an increase in the use of partial replacement despite on-going concerns about the higher rates of failure. Currently there is no clear consensus within the orthopaedic community about when a unicondylar or total knee replacement should be used. Surgical practice varies dependent upon the distribution of osteoarthritis, the severity of pre-operative symptoms (Lutzner 2009), patient age (Mancuso 1996), obesity (Cobos 2010) and the surgeon's own preference.

For the majority of surgeons total knee replacement remains the 'gold standard' procedure for surgically treating knee osteoarthritis. Greater than 90% of all knee replacements performed in England and Wales in 2011 were total knee replacements (72,126 of 79,516) (NJR-AR 2012). In contrast only 9% of knee replacements were partial knee replacements (7,390 of 79,516 (1,113 Patello-femoral, 6,257 unicondylar)). These proportions have been consistent over the last 7 years (NJR-AR 2012) (Figure 1.4).



Year	2005	2006	2007	2008	2009	2010	2011
TKR using cement	84%	83%	83%	83%	83%	85%	86%
TKR not using cement	7%	7%	6%	6%	6%	5%	4%
TKR Hybrid	1%	1%	1%	1%	1%	<1%	<1%
Patello-femoral	1%	1%	1%	1%	1%	1%	1%
Unicondylar	7%	8%	8%	8%	8%	8%	8%
Number of procedures	59,385	62,327	73,523	77,624	78,832	81,042	79,516

Figure 1.4: Types of knee replacements performed in England and Wales 2005 to 2011 (Taken from NJR 9th Annual Report 2012).

1.2 National arthroplasty registers

1.2.1 The evolution of knee arthroplasty registers

The first national knee arthroplasty register commenced in Sweden in 1975 (Lidgren 2009). Its primary aim was to give early warning of inferior implant designs and present the average results for knee replacement based on the experience of a whole nation (Robertsson 2007). Data from this registry have been used to quantify implant failure rates for several decades in the Swedish population (Knutson 1986, Lidgren 2009). Following the success of this initial register many countries developed national registries with the aim of reducing implant revision rates, thereby reducing morbidity, mortality and costs associated with further surgical intervention (Havelin 1994, Graves 2004). Worldwide there are now more than 20 registers either actively collecting data or being piloted (Bohler 2009) and, because of their success, there has recently been a combined call for an increase in the number of registries from the presidents of a number of national orthopaedic associations (Pellegrini 2009).

On the surface registries are primarily health technology audits of the procedures that have been performed in a given country (Robertsson 2007). They prospectively follow-up all patients with no limits on who is included and what implants are used. By nationally monitoring patients who undergo knee replacement they are able to assess the relative success of different approaches to surgery, different implants and also the results of individual surgeons and centres. They reflect the experience of a whole nation rather than the specific results of an individual surgeon or centre. As such they better represent the results that can be expected following knee replacement procedures when compared to information derived from case series, isolated cohort studies or prescriptive randomised trials (Robertsson 2007).

Bohler identifies three key aims that are fundamental to any successful registry. Firstly registers must record all primary and revision operations in a defined area in a central

database. Secondly they should follow each primary implant until it has to be revised, the patient dies or emigrates. Thirdly, they need to use a consistent endpoint, usually revision, which should be defined in a standardised manner (Bohler 2009). The standardised definition of revision removes ambiguity for the clinicians coding and gathering the data and helps to ensure data accuracy in relation to the registries primary outcome measure. For most registries this definition is the removal or exchange of at least one part of the original knee implant (Bohler 2009).

A number of key elements have also been suggested as necessary for the successful development of a national registry (Bohler 2009). These include:

- Integration of the registry within a national health care system.
- A professional, central structure for data collection and storage.
- Understanding of national issues regarding data protection of personal data.
- Involvement of specialist experts for the interpretation and statistical analysis of data.
- The close involvement and support of professional orthopaedic associations within the host country.

1.2.2 The role of knee arthroplasty registers

There are a number of benefits to knee replacement registers. They report national trends in patient demographics, implant usage and surgical methods which can help with healthcare monitoring, purchasing and future provision planning. They allow for the early identification of implants, surgeons and centres associated with higher than expected levels of failure. Similarly newly introduced implants and surgical techniques can be rapidly and accurately evaluated. Monitoring performance may also be a deterrent for the use of implants and methods that have not been formally evaluated. Publication of results acts as a national benchmark for performance against which comparison can be made. It also provides information on the best surgical techniques and expected outcomes and helps

guide implant selection. National recording of operations also permits for easy identification and recall of implants in the case of product recall or the failure of a specific implant. This information not only benefits patients but also help to limit health care costs by reducing the burden of revision surgery.

National sampling is not without its difficulties. In addition to the obvious logistical issues there are on-going questions surrounding the completeness of data within registries, compliance with data collection and the quality of data recorded. In large multicentre registers there is an inverse relationship between the amount of data collected and the quality of data delivered (Robertsson 2007). There is therefore a balance to be struck between ambitious data collection and the completeness and accuracy of the data on which registers rely. Most registers therefore currently collect a limited dataset to try and ensure the completeness of the data collected (Robertsson 2000). There is currently a drive to standardise this 'minimal' dataset to allow better comparison and collaboration between worldwide registries (Robertsson 2007).

Another issue is how registries measure outcome. The strengths of national sampling and broad inclusions produce problems due to the variations in surgical routines and follow-up employed in the different centres submitting data. This method of data collection is different to those employed in large scale clinical trials which utilise strict inclusion criteria and standardised follow-up (Robertsson 2007). A lack of standardised follow-up mean registers are limited in the type of outcome they can assess. Radiological and functional outcomes assessed at differing time points are not comparable and therefore not suitable as a method of outcome. There is, however, one outcome pertinent to knee replacement that provides a definite measurable outcome; revision surgery. Revision brings the patient back into the healthcare system and provides an event that is recordable and linkable to previous surgical episodes.

This measure is, however, not without its limitations. Revision is crude measure of outcome. It does not account for knee replacements that would be deemed as failures due to poor function and on-going pain that do not undergo further surgery (Murray 1993, Wylde 2011). In addition certain patients may not be fit to undergo revision or may not consent to surgery even in the face of a failing implant. These patients will be overlooked using this method of assessment. Furthermore patients revised outside their original healthcare system may not be recognised leading to an underestimation of the rates of revision.

Registries fulfil different roles within the healthcare systems in which they have been developed and this is reflected in the way they present and publish their data. Some simply measure implant, centre and surgeon performance using revision as their sole outcome measure (England and Wales (NJR-AR 2012), Australia (AJR-AR 2012). Their annual reports are dominated by comparisons of the rates of revision and mortality for different implants and approaches to surgery. The Scandinavian registries (Sweden, Norway, Denmark (NARA 2012)) have progressed this idea further publishing rates of revision and mortality for individual regions and centres and making this information available to the public. This approach is also seen in the Scottish registry where revision rates in addition to information on patient mortality, morbidity and post-operative complications are reported at a centre level (SAP-AR 2011). This kind of reporting aims to improve quality by making results for each centre publically available. More recently both the Danish arthroplasty register and the Swedish hip and knee arthroplasty register have started to record and report patient reported functional outcomes and satisfaction data as part of their routine collection. The Danish registry has perhaps been the most progressive in this respect as it routinely reports on six quality indicators which are published at a centre level (Danish arthroplasty register AR 2011). These indicators include:

1) Rate of intraoperative complications – complication rate should be <10%.
2) Rate of Follow-up (important as it allows follow-up PROMs collection) – follow-up rate should be >90%.
3) Levels of satisfaction (Collected at 6-18 months post-operatively) – >90% patients should be very satisfied or satisfied.

4) Improvement in knee function (Using the Knee Society Score (KSS) collected at 6-18 months) – >90% patients should improve by ≥25 points.
5) Implant survival (adjusted for age and sex) – Reference provided for each centre against the population mean.
6) Mortality – Low mortality is deemed to represent high quality but no reference point given.

This level of reporting provides a benchmark against which all centres can be compared and allows transparency of reporting by putting this information in the public domain.

A summary of the current key worldwide national registers is given in table 1.1. This demonstrates the variation in outcome reporting for the 10 registries for which results were publically available in the English language. It also reveals the disparity in data collection which covers only 42% of knee replacement procedures in Canada but can approach 100% for other registries (Australia, England and Wales, New Zealand, Scotland). For a further 12 registers data was not accessible at the present time.

Registry (Annual Report (AR) reviewed)	Originated	Number of knee records (up to Year)	Coverage / Compliance with data collection	5 year survival reported	10 year survival reported	PROMs reported	Satisfaction reported	Complications (Intraoperative or Postoperative) reported	Mortality reported	Survival stratified by reason for revision reported	Revision outcomes (Survival/PROMs) reported
Australia (AR 2011)	1999	333,764 (2010)	"Almost complete dataset"	Yes	Yes	No	No	No	No	Yes	Yes (re-revision rates)
Canada (AR 2010)	2001	124,783 (2010)	42% (2010)	No	No	No	No	No	No	No	No
Denmark (AR 2010)	1997	60,049 (2010)	92% (2009)	Yes ¥	Yes ¥	Yes ¥ (KSS)	Yes ¥	Intraoperative - Yes ¥ Postoperative - Yes ¥	Yes ¥	No	No
England and Wales (AR 2011)	2003	>500,000 (2010)	Approximately 100% (2009-11)	Yes	No	No	No	Intraoperative - Yes Postoperative - No	Yes	No	No
Portugal (AR 2010)	2009	4,401 (2010)	59% (2010)	No	No	No	No	No	No	No	No
New Zealand (AR 2010)	1999	62,408 (2010)	98% (2009)	Yes	Yes	Yes	No	No	No	Yes	Yes (Re-revision rates)
Norway (AR 2010)	1987 (1994 (for Knees)	41,542 (2009)	>95% (2009)	Yes	Yes	No	No	No	No	No	No
Scotland (AR 2010)	2001	54,717 (2009)	99% (2010)	Yes #	No	No	No	Intraoperative - Yes # Postoperative - Yes #	Yes #	No	No
Slovakia (AR 2010)	2003 (2006 (for knees))	8,385 (2010)	N/A	No	No	No	No	No	No	No	No
Sweden (AR 2011)	1975	175,345 (2010)	96.6% (2009)	Yes #¥	Yes #¥	Yes (Piloted in 2011 report - KOOS, EQ5D, Pain VAS)	Yes (Piloted in 2011 report)	No	No	No	No
Others: Currently the following countries also have arthroplasty registers (From www.ear.efort.org/registers.aspx): Austria, Belgium, Croatia, Czech Republic, Finland, France, Hungary, Germany, Romania, Slovenia, Switzerland, Malawi Data is not currently available from these registers for the following reasons 1) No results as yet reported 2) No public access to reports 3) reports no published in English language.											

Table 1.1: Summary of the current key national knee registers. Where available, data is presented on the currently reported outcomes for each registry.,

Key: N/A = Not available, ¥ = Outcome reported for individual centres, # = Outcome reported for individual regions, KOOS = Knee injury and Osteoarthritis Outcome Score, EQ5D = Euroqol 5D score, VAS = Visual Analogue Scale.

1.3 The National Joint Registry for England and Wales (NJR)

1.3.1 History of the NJR

Sir John Charnley first suggested the idea of a national arthroplasty register in the 1970's at a time when joint replacement was becoming increasingly popular. He recognised that to ensure high standards of care the surgical practices and implants in everyday use needed to be monitored and audited. However, despite this it was a number of years before the first regional registries were established in the Trent region (Fender 2000) and later the north west of England. Publications from these regional registries maintained the impetus for a national register but it was only after the Royal College of Surgeons report into the failure of the 3M capital hip (Massoud 1997), alongside recommendations form the National Audit Office and the National Institute for Clinical Excellence that the National Joint Registry for England and Wales (NJR) was finally established.

The NJR was established by the Department of Health in 2002 and started collecting information on hip and knee replacements performed in England and Wales in April 2003. It is currently operated by the Healthcare Quality Improvement Programme (HQIP) and is the largest joint registry in the world. Costs are met by a levy of £20 (2010/11) raised on every acetabular, talar and distal femoral component sold. The NJR is a keystone to delivering the commitment of both the Department of Health and the Welsh Assembly Government to improve the health and wellbeing of the population and is a vital tool for improving national clinical standards for hip and knee replacements.

The NJR collects data via a web-based system with surgeons providing information on the characteristics of the patients, the implants, surgical techniques, and the need for revision or re-operation. In March 2012 the total number of procedures recorded in the NJR passed 1.2 million with > 99% of all hospitals on the NJR database submitting data (NJR-AR 2012). By linking patients within the registry, revision of a previously implanted primary can be identified and a cause of failure assigned. The data can be used to describe trends over

time, such as an increase in the mean body mass index of TKR patients, or identify implants or surgeons with significantly elevated revision rates—'outliers.'

The NJR measures its performance against 3 key indicators: compliance, consent and linkability. Compliance is measured by comparing the number of procedure forms submitted to the NJR against the number of levies through implant sales and was 90.3% for 2011/12. The consent rate is determined by comparing the number of records where the patient has agreed to their personal data being stored on the NJR database compared with the total number of procedures recorded on the NJR. The rate of consent was 90.4% for 2011/12. Linkability looks at the number of records for which an NHS number was entered compared with the total number of procedures recorded on the NJR. This is important as the NHS number is required to link all primary and revision procedures relating to a single patient. The rate for 2011/12 was 95.5%.

Despite high levels of compliance, consent and linkability little is currently known about the quality of data held within the NJR, specifically whether the data recorded on the database is representative of the clinical activity undertaken. Some work is currently being undertaken by the NJR to address this by looking at the quality of revision coding in the NJR database compared to another national dataset, the hospital episode statistics (HES). Discrepancies in revision coding between these datasets will be identified and a reason determined. It is hoped this will give some insight into the quality of revision data held within the NJR. However, despite this initiative more clearly needs to be done to ascertain whether the data stored and used by the NJR is truly representative.

1.3.2 Aims of the NJR

At the heart of the NJR is a database of information collected from all the hip and knee replacement procedures in England and Wales. Using this data the NJR aims to:

- *Ensure patients obtain the best clinical care during and following their joint replacement operation.*
- *Ensure that NHS and other healthcare resources are best used*
- *Improve surgical practice through the identification of best practice in orthopaedic units/ hospitals.*
- *Highlight in real time any brand of prosthesis showing high failure rates and allow prompt removal from the market, if necessary.*
- *Improve evidence based purchasing of joint replacement implants for orthopaedic units/ hospitals.*
- *Provide patients, clinicians, healthcare purchasers / commissioners, regulators and implant suppliers with evidence for which are the best performing implants.*

(Taken from www.njrcentre.org.uk)

1.3.3 The NJR in 2013

With over 1.2 million records the NJR is now the largest arthroplasty register in the world. In the last few years the NJR has taken a number of strides to increase the quality of its data and the outcomes it is able to report. In April 2011, the Department of Health made participation in the NJR mandatory for all NHS hospitals in England and the Welsh Government agreed mandation for Wales. This will help to ensure data completeness for primary hip and knee replacements being undertaken in England and Wales. Its sub-committees have worked alongside implants manufacturers, the Medicines and Healthcare products Regulatory Agency (MHRA), and individual trusts and surgeons to identify performance outliers. The research subcommittee has been working to streamline the process of accessing NJR data for the purposes of research and increasing research output.

In 2011 the NJR started collecting data on ankle replacements to supplement the data already collection for hip and knee replacements. By 2013 it should also be collecting data on shoulder and elbow replacements. There are also plans to increase the geographical

scope of the NJR with Northern Ireland likely to become included in the registry over the next few years. The NJR is also currently involved with the Department of Health (DoH) in the collection of patient reported outcome measure data enabling them to deliver even richer data for analysis, study and research.

1.4 The current status of registry research: Where does it fit in?

1.4.1 Research into knee replacements: Where are we in 2013?

Over the last 20 years there has been a progressive movement toward the use of 'evidence' to support clinical decision making (Sackett 2000). Initially evidence based medicine (EBM) and latterly evidence based practice (EBP) and evidence based treatment (EBT) have developed to provide clinicians with information about the effectiveness of clinical interventions. While there has been widespread adoption of these processes there remains an on-going debate about the nature of evidence and particularly how evidence supports the use of therapeutic interventions (Rawlins 2008).

Presently hierarchies are used to 'weight' evidence with evidence obtained from randomised trails ranked higher than evidence derived from observational sources such as cohort, case-control and cross sectional studies (Petrie 2006). These rankings are based on the reliability and strength of the evidence, and its ability to determine causation (Rawlins 2008). The debate about the nature of evidence has connotations for the way the current body of evidence is viewed and how research into knee replacement is performed in the future.

Supporting evidence for knee replacements has traditionally originated from small single-surgeon or single-centre observational case series (Carr 2012). Reports on new knee replacements were often from the personal series of surgeons involved in the design of these implants and were therefore prone to bias and potential conflicts of interest (Font-Rodriguez 1997, Pandit 2011). These papers reported uniformly good results and led to a proliferation of implants available to the surgeon (NJR-AR 2012) many of which are being used without evidence of their clinical effectiveness. Randomised trials comparing implants were rarely performed to established clinical effectiveness.

Due to this proliferation the orthopaedic community is left in the unenviable position of having a large number of knee replacements in current use without any good evidence on the comparative effectiveness of each implant or approaches to surgery that might help guide clinical decision making. The use of widespread observational research has meant knee replacement surgery is mired in lower level evidence that lacks the ability to make strong recommendations about what should be considered best practice. Formal clinical evaluation requires large numbers of patients and long-term follow-up to adequately assess the primary safety end point, which is long-term device failure or need for revision (Bohler 2009). Furthermore assessment of the primary effectiveness end-points which are short and longer term patient reported outcomes require prospective data collection, planning, and substantial funding which present logistic and financial problems that are difficult to overcome with a clinical trial design (Sedrakyan 2011).

Randomised controlled trials could help as they provide the best evidence of a causal relationship, and thus the strongest evidence on which to base practice. However, obtaining evidence in this way is likely to be extremely difficult. Randomised trials are costly, logistically complex and time-consuming. Given the large number of comparators that would need to be assessed and the fact that many of the outcomes of interest are relatively rare events it would seem that the size and the scope of such studies would make them unfeasible. For example, a trial designed to compare the relative rates of failure of 2 interventions would require approximately 1000 patients in each study arm to detect a 50% reduction in failure from 5% to 2.5% with a power of 0.80. The numbers of patients required to detect smaller differences in the rate of failure or to make comparison between more than two groups would require significantly larger numbers. The numbers involved in even the simplest trial are therefore likely to be prohibitive to its completion.

1.4.2 Can registries help?

One option to help provide evidence to support decision making is re-evaluate the role of observational data, and specifically to use the data held within national joint replacement registries. National registries compliment more traditional research designs (e.g. randomised controlled trails (RCTs)) by providing important observational data on the effects of specific treatment strategies within national clinical practice (Kolling 2007). They allow the outcome of care to be linked to current health care practices. Due to the large number of patients available and the extended period of observation within these databases they are well suited to detecting rare events and differences between interventions in the longer term (Garellick 2000).

National observational studies have obvious advantages when compared to a randomised controlled trial. They contain a large number of patients giving high statistical power, they make it possible to perform adequate analyses of uncommon complications, and they limit the effects of performance bias (NARA 2012). Furthermore they can often be completed in a short space of time, they do not infer significant costs, are logistically easier than randomised trials to instigates and complete, they provide information on a large number of decisions, interventions and outcomes, and they provide 'real world' information. As they do not discriminate or exclude they often include groups of patients often omitted form randomised trials. However, because their primary function is not the production of research they often contain limited data and the research questions it is possible to answer are constrained by the data available. Importantly these databases rarely include information about clinical decision making; why does the surgeon choose to pursue one intervention in one patient on one day and an alternative intervention in a different patient the following day. This is important as the patient characteristics that inform clinical decision making may also influence clinical outcomes. This leads to uncertainty about whether it is the intervention or the myriad of patient factors that causes any observed differences in outcomes. It can therefore be difficult to attribute causality using observational registry data, a pitfall that must be appreciated.

The problem of attributing causality seen with observational data is less of an issue for randomised trials. This is because the data to be collected is determined as part of the research question and trial design and surgical decision making is abolished by randomisation, as long as the patient meets the relevant inclusion criteria. Randomisation eliminates a lot of the uncertainty that troubles observational research. However it is not without its problems, particularly in the evaluation of knee replacement outcomes. Randomised trials are designed to answer specific research questions and are not suited to the analysis of epidemiology and demography which can be assessed using registry data. The predetermined limitations imposed by trial design hamper their ability to identify additional factors that may potentially influence outcome (Graves 2010). Unfortunately many randomised trials also ask the wrong questions. Clinicians, policy makers and patients want to know which intervention works the best?, is intervention 'A' better than intervention 'B'?. However, many randomised trials simply ask whether a specific intervention works by comparing the intervention in question to a placebo when what clinicians really want to know is how an intervention compares to the current 'gold standard' best practice. Unless a pertinent research question is posed there is a risk that the randomised trial is reduced to nothing more than an efficacy study which simply establishes if an intervention can work, rather than a clinical trial which establishes which intervention is best. This information may not be helpful to the clinical decision making process and is often removed from the conditions that clinicians face in daily practice.

Randomised trials also suffer from problems of generalisability. In the knee replacement setting trials undertaken by experienced enthusiasts for a given procedure on an artificially defined subset of patients may not be representative of what happens in the average clinician's daily practice. This limits the validity of the trials conclusions. In addition trials have defined start and end points and cannot comment upon actions outside this timeframe. Given these limitations it is questionable whether the randomised controlled trial can ever replace registry analysis in the assessment of longer term outcomes following knee replacement. For most replacements the time required to observe even a small number of failures mean the number of patients that would need to be enrolled and the

duration of follow-up for any prospective comparative trial would be prohibitive to its completion.

A review of current nationally funded clinical trials put these logistical problems in context. In the last 10 years the National Institute for Health Research's (NIHR) health technology assessment (HTA) programme has only funded two randomised clinical trials directly comparing different approaches to knee replacement. The Knee Arthroplasty Trial (KAT) (HTA ref: 95/10/01) commenced in 1998. It is a randomised controlled trial of different knee prostheses that aims to answer 4 main questions: 1) Should the inner surface of the knee cap be resurfaced? 2) Should the tibial component have a metal back? 3) Should the knee replacement have a mobile bearing? 4) Should unicompartmental or total knee replacement be performed?. Outcome data collected includes information about short-term complications, quality of life, functional outcomes and cost-effectiveness. The project is due to be completed in November 2013, 15 years after is commenced, and current HTA funding is £1,381,371. The Total or Partial Knee Arthroplasty Trial (TOPKAT) (HTA ref: 08/14/08) aims to "assess the clinical and cost effectiveness of Total Knee Replacements *versus* Unicompartmental Knee Replacements in patients with medial osteoarthritis". It commenced in 2010, with results expected in early 2020. HTA funding for this project is £2,700,878. These two trials highlight the enormous cost (>£4 million), resource implications and time (10-15 years) necessary to undertake randomised clinical trials in knee replacement surgery.

1.4.3 The way forward?

In his recent Harveian oration Prof Sir Michael Rawlins, Chairperson of the National Institute for Clinical Evidence (NICE), stated *'The notion that evidence can be reliably placed in hierarchies is illusory. Hierarchies place randomised controlled trials on an uncomfortable pedestal as although they have advantages they also have disadvantages'* (Rawlins 2008). Furthermore he states that *'Decision makers need to assess and appraise all the available*

evidence irrespective of whether it has been derived from randomised controlled trials or observational studies; and the strengths and weaknesses of each need to be understood if reasonable and reliable conclusions are to be drawn' (Rawlins 2008).

Once the relative merits and pitfalls of both observational and randomised trial data are understood it becomes clear that neither approach is perfect, but instead it suggests that these two approaches can provide complementary information. This appreciation has led to the suggestion of an alternative research approach: Comparative effectiveness research (Sox 2009). Comparative effectiveness research is defined as:

'The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. The purpose of comparative effectiveness research is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels'.

Within this definition is the intention to directly compare the effectiveness of interventions in patients who are typical of day-to-day clinical care. Furthermore this approach allows for the identification of clinical characteristics that predict which interventions would be most successful in individual groups of patients so that interventions can be tailored to different sub groups of patients. This emphasises the increasing appreciation that pragmatic trials often represent an artificial situation removed from general clinical practice and that research needs to provide the information that decision makers need to know. It understands that randomised trials are also constrained by the research questions they pose limiting the ability to make potentially important inferences about specific sub groups and alternative interventions. This approach has recently received support in the United States with a funding award of $1.1 billion from congress and presidential support (Sox 2009).

Comparative effectiveness research appreciates the limitations of both observational research and randomised trials discussed previously. The committee charged with delivering comparative effectiveness research has implicitly stated that '*Very large collections of the electronic records of patients can be a valuable resource for CER*'. They believe that '*By using these data sets, it is possible to compare the outcomes of several effective interventions in a population that is representative of daily care. The great numbers of patients in these data sets also makes it possible to study subgroups with precision and perhaps identify key predictors of response to an intervention, both of which would facilitate decision making at the individual and population level. These are features that are aligned with the goals of CER*' (Sox 2009). This has led to a recommendation that '*The CER program should help to develop large scale clinical and administrative data networks to facilitate better use of data and more efficient ways to collect data to yield CER findings'* (Sox 2009).

1.4.4 Current research output from registries

The comments from Prof Sir Michael Rawlins and the CER committee suggest an increasing role for registry analyses in the formation of clinical evidence. In a recent review of the scientific production and impact of national registers Boyer *et al* (2011) concluded that compared to randomised controlled trials registry papers are less often published but are more often cited. This gives credence to the idea that registry analysis is becoming increasingly accepted as a source of evidence in modern orthopaedic practice.

Boyer's analysis found that between January 1st 1980 and December 31st 2008 there had been 190 registry publications from a total of thirteen national arthroplasty registers (Swedish Hip, Swedish Knee, Finnish, Norwegian, Danish hip, Danish Knee, Australian, New Zealand, Scottish, Canadian, Romanian, England and Wales, Slovakian). January 1st 1980 was chosen as the start point for their analysis as the first article published from a national register was published in 1980 (Ahnfelt 1980). Over the same time period they identified

476 randomised controlled trials and 40 meta-analyses that were published relating to either hip or knee replacement surgery. The mean journal impact factor for randomised trials was 1.9 [IQR 1.6 to 2.2] compared to 1.8 [IQR 1.8 to 2.2] for registry publications. The total citations and 3 year citations for randomised trials were 7.0 [IQR 2.0 to 20.0] and 2.0 [IQR 1.0 to 6.0] compared to 13.0 [IQR 2.0 to 31.0] and 3.5 [IQR 1.0 to 6.0] for registry publications (Boyer 2011). This demonstrates that the higher rates of citations for registry papers came from journals with equivalent scientific impact.

1.5 Introduction summary

Establishing best practice for knee replacement is important given the large number of knee replacements performed worldwide. Knee replacement research is problematic given the large number of implants and procedures being performed and the logistical and financial limitations associated with prospective clinical trials. Observational studies and clinical trials have different strengths and weaknesses. For the evaluation of knee replacements these approaches should be viewed as complementary, with an appreciation that for longer term outcomes that occur infrequently registries may be the best model for assessment. Registry data offers a “real-time” perspective of national practice that can be used to answer clinically relevant research questions. Previous publications from registries are well cited and achieve publication in journals with comparable scientific impact to randomised trials. This suggests that registry analysis have an increasing role to play in the generation of orthopaedic evidence.

Chapter 2: Aims

2.1 Thesis aims

Observational registry research clearly has an increasing role to play in the production of evidence to guide decision making surrounding knee replacement. There is currently a significant output from registries relating to knee replacement outcomes. However, the output from the NJR is poor and with over 1.2 million operative records it represents a wealth of clinical data that is currently underexploited. Furthermore it remains unclear if observational registry data is able to answer the same clinical questions posed by randomised clinical trials, how much these two approaches overlap and how much additional value registry data provides.

This work therefore has the following aims:

1. To assess whether research performed on data recorded by the NJR has the ability to answer clinically relevant research questions relating to knee replacement surgery.
2. To determine if registry research is able to answer specific clinical questions that are unsuited to prospective randomised clinical trial designs.

These broad aims will be achieved by identifying pertinent clinical research questions covering differing aspects of knee replacement practice. Attempts will then be made to answer these questions using the registry data available. Subsequently, the conclusions drawn and the strengths and weaknesses of any analyses will be discussed with particular reference the methodological differences between registry analyses and prospective randomised clinical trials (specifically KAT and TOPKAT). To achieve these aims this work will focus on three complimentary areas of research which will be addressed individually in three results chapters (chapter 4, 5 and 6). These are:

- Chapter 4: Clinical outcomes after unicondylar knee replacement (UKR).
- Chapter 5: Clinical outcomes after total knee replacement (TKR).
- Chapter 6: Clinical outcome after revision total knee replacement (RTKR).

Chapter 3: Methods

3.1 Systematic review of current registry publications relating specifically to knee replacement procedures

As previously discussed the scientific production and impact of publications from national registers has recently been reviewed by Boyer *et al* (2011). However, due to the search criteria used in this analysis (inclusion of both hip and knee replacements publications and inclusions of both clinical and non-clinical (papers on registry development, organisation, statistical methods and reporting observational data similar to that found in annual reports) publications) it was unclear exactly how many of these publications related specifically to clinical research on knee replacements.

A further literature search was therefore conducted to establish the current number of registry publications relating to outcomes after knee replacement procedures and specifically the number of publications from our own national registry, the National Joint Registry for England and Wales (NJR). The literature search was performed on the 9th March 2011 using Medline (1946 to 2012), Ovid Embase (1974 to 2012), and Web of knowledge (limited to English language, research articles and orthopaedic subject area). The search terms used within each database are given in table 1.2. Using this search strategy a total of 751 papers were identified.

Abstracts for these papers were then screened and excluded if they fulfilled any of the following criteria:

1) From an institutional registry/database i.e. not a national/regional registry
2) Paper was not a clinical research paper e.g. review paper or commentary on registries, papers describing the developments, organisation, statistical methods of registries, reporting of observational data similar to that found in annual reports
3) Conference abstracts or proceedings
4) Not written in the English language

5) Articles relating to primary knee arthrodesis
6) Articles relating to other joint arthroplasty (e.g. Hip, ankle)
7) Articles relating to soft tissue knee procedures (e.g. Anterior cruciate ligament repair)
8) Duplications

Database	Search terms	Number of articles
Medline (1946 to 2012)	MeSH terms 1. "Registries/" 2. "Knee prosthesis/" *or* "Osteoarthritis, Knee/" *or* "Knee/" *or* "Arthroplasty, Replacement, Knee/" *or* "Knee Joint/" 3. "1" and "2"	43,514 57,127 **184**
Ovid Embase 1974 to 2012	4. "Registry.mp" or "register" 5. "Knee arthroplasty" *or* "Knee prosthesis" *or* "Knee osteoarthritis" *or* "Total knee replacement" 6. "4" and "5"	81,033 31,295 **262**
Web of Knowledge	7. "Knee" and "registry" *or* "Knee" and "register"	**305**
TOTAL		751

Table 3.1: Search criteria and numbers of hits for knee replacement literature search.

Where any doubt remained over the content of the paper it remained included. This initial screening left a total of 94 papers. Additional papers were then sought by reviewing the relevant research sections of all of the previously mentioned national registries (Table 3.1). This produced an additional 38 papers giving 132 in total. These papers were then reviewed with further exclusions based on the criteria given above (Table 3.2). Final paper count was 59 and included only two papers from the NJR (Sibanda 2006, Baker 2007). The majority of these papers originate from the Scandinavian registries (Sweden/Norway/Denmark) and

focus on implant survival as the primary outcome. The number of publications from the NJR is small given its size relative to other worldwide registries (see table 1.1) the majority of which have previously been used to support a wide range of clinical research projects.

Reason for exclusion	Number of articles
Start	132
Not national registry papers - From a single institution - From a community or regional registry - From a national administrative database other than a national arthroplasty register	 4 18 3
Papers describing development, organisation, statistical methods of registries. Reporting of observational data similar to that found in annual reports	17
Review paper or registry commentary	18
Papers describing statistical analysis of registry data	6
Conference abstracts	5
Non English language	1
Hip paper	1
Total remaining after exclusions	**59**

Table 3.2: Details of the exclusions used in determining the currently published clinical research from national knee arthroplasty registers.

3.2 The NJR Research process

3.2.1 The NJR research strategy

"The NJR aims to provide a substrate for definitive research into the full range of biological, mechanical, clinical and social factors influencing the outcome of joint replacement and to establish the impact of joint replacement surgery on the well-being of patients and the population. The aim is to enhance the understanding of the science of arthroplasty, improve and enhance clinical practice and benefit public health. The NJR aims to provide extensive links to other population health resources and to encourage the widest possible access of the data to the research community through providing a platform for enquiry into all aspects of arthroplasty." (www.njrcentre.org.uk)

3.2.2 The research sub-committee

The research sub-committee oversees all research undertaken using NJR data. It is directly responsible to the NJR steering committee. Their stated aim is to *"maximise the value of the NJR to research by making data widely available."* The 7 person committee is chaired by Professor Alex MacGregor and includes 4 consultant orthopaedic surgeons, 1 orthopaedic industry representative and 1 representative from the healthcare quality improvement programme (HQIP). The committee oversees the research strategy of the registry as well as reviewing and approving all of the research requests submitted. It takes responsibility for the release of data for research through an *"impartial and objective*" protocol and it is committed to *"upholding the standard and consistency of work that is carried out on NJR data."* The Research sub-committee supports the national strategy for orthopaedic research. Within the NJR its specific remit is (from www.njrcentre.org.uk):

- To act as a point of contact to provide impartial and informed advice on planned research using NJR data.

- To review all research data requests received on behalf of the NJR using an established protocol and manage the release of research data.
- To confirm that proposed research is scientifically appropriate, methodologically feasible and is an appropriate use of the resource.
- To confirm that proposed research conforms to current ethical standards.
- To monitor the use of NJR data through an established reporting procedure.
- To review all abstracts, presentations and publications arising from the NJR data.
- To communicate research activity through the NJR website, NJR annual report, and the wider media.
- To provide quarterly reports to the NJR steering committee.

3.2.3 The NJR research application process

Prior to 2010 data requests were made using a web based request form accessed via the NJR website. Applications took a substantial time to process (up to 6 months) due to a lack of a co-ordinated approach between the application process and the research sub-committee. As the volume of research requests increased it was realised that this process was failing researchers and so in June 2010 the NJR appointed a research officer to co-ordinate research being undertaken by groups external to the NJR. At the same time a new research process was developed in an attempt to streamline the application process. This involved the production of a standardised data request form (Appendix i), downloadable from the research area of the NJR website, which, when completed, is sent directly to the research officer. The research officer then directly circulates the requests to members of the research sub-committee for approval with a decision expected within 3 weeks. All applications for data made as part of this analysis were undertaken using this process.

Once approved, the data requests are forwarded to the NJR data handlers (Northgate Information Solutions) for NJR data, or the Department of Health data handlers (NHS Information Centre) for PROMs data. Following satisfactory completion of relevant

information governance, data handling and data re-use agreements the data can then be released to the requester.

3.2.4 Data applications for the current work

Data for the current analysis were accessed in accordance with the standard NJR procedure relating to data access described above. This involved three separate applications for NJR data, one of which included a request for linked PROMs data. These included:

- NJR extract: Analysis of revision knee arthroplasty in England and Wales. Approved by the NJR research sub-committee on the 6th August 2010. Data received 29th August 2010.
- NJR extract: Examination of revision rate following knee arthroplasty by indication for revision. Approved by the research sub-committee on the 3rd May 2011. Data received 27th July 2011.
- NJR-PROMS linked extract: PROMs following primary total knee replacement and unicondylar knee replacement. Approved by the NJR committee on the 6th August 2010 and approved by the NHS information centre on the 13th December 2010 (Ref NIC-74514-LPDX4). Linked data received on the 26th April 2011.

Following preliminary analysis of the information contained within these data requests subsidiary projects were developed that fell outside the remit of the initial research requests. However, these projects did not require any additional data over and above that already provided. For these projects a dialogue was opened with the NJR research officer and the chair of the NJR research committee by which these additional projects could be undertaken as long as they were logged with the research committee on a project by project basis. In addition all work was subject to internal NJR review prior to release for publication or presentation in the public domain.

3.3 Administrative considerations

3.3.1 Information governance

Prior to accessing the NJR and PROMs data the National Information Governance Board for Health and Social Care (NGIB) ethical and confidentiality committee was contacted to discuss the need to gain permission under section 251 of the NHS Act 2006 to access NJR data for research. Section 251 allows the common law duty of confidentiality to be set aside in specific circumstances where anonymised information is not sufficient and where patient consent is not practicable. After discussion it was decided that section 251 would not be required for any of the proposed projects as long as patient identifiable data was not requested or transferred. To overcome this issue data requests were designed so that any patient identifiable data (name, date of birth, address, post code etc.) was either removed or anonymised at source prior to transfer.

Data was stored on a secure server at Northgate Information Solutions (the NJR data handler). This server was then accessed remotely using encrypted laptops provided Northgate and conforming to their own information security standards. The NJR also permitted some subsets of NJR only data (i.e. not PROMs data) to be stored on local NHS and/or University servers.

3.3.2 Caldicott approval

Data handling, storage and security were performed in accordance with local trust policies on advisement of the local Caldicott guardian. Caldicott approvals were obtained via Richard Oliver the Newcastle NHS trust Caldicott guardian for the following projects:

- Analysis Of Revision Knee Surgery Practice Within England And Wales Analysis Of Data From The National Joint Registry For England And Wales (Caldicott number 1220).

- Patients Reported Outcomes Measures proms Following Primary Total Knee Replacement Vs Unicondylar Knee Replacement (Caldicott number: 1280).

The Caldicott guardian for additional data requests was Elaine Young the Healthcare Quality Improvement Programme (HQIP) lead for the NJR on behalf of Northgate Information Solutions the NJR data handler.

3.3.3 Ethical considerations

Prospective projects were discussed with the chairs of the local ethics and research and development committees. Correspondence with Paddy Stephenson chair of the Sunderland local research and ethic committee can be found the appendices. As no patient identifiable data was requested and as no additional patient contact was required it was decided that all projects could be performed as service evaluations without the need for additional ethical approval. The NJR has its own consent mechanism for data collection and it was felt that this was sufficient for the purposes of the proposed research.

3:4 Overview of the databases available for the current analysis

3.4.1 Introduction

The primary aim of this work was to assess the ability of the NJR to answer clinically relevant research questions relating to knee replacement surgery. As such the data held within the NJR were chosen as the basis for our analysis. However, we were aware that when used in isolation the outcome measures available from registry data are limited. This is because the only outcome measures available within the NJR are implant revision and mortality. As discussed previously implant revision is a crude measure of failure and does not differentiate between patients performing well and those living with a painful replacement. This methodology has therefore been cited as a major limitation of registry data (Wylde 2011).

To provide further information on implant performance we decided to link the NJR database to the patient reported outcome measures (PROMs) programme instigated as a joint venture by the Department of Health (DoH) and the NJR in 2009. The PROMs programme addresses some of the problems encountered when looking at registry data in isolation by providing validated subjective and objective measures of patient health and disability, both prior to and at specific intervals following surgery (PROMs 2011). Linking these two large, population-based databases provides the capability to analyse both functional and clinical outcomes and allows for the analysis of a wider range of variables that could potentially influence outcome.

3.4.2 Overview of the data available from the National Joint Registry (NJR) for England and Wales database

The NJR has assimilated data on patients, surgeons and implants performed in both the private and public sector (National Health Service) in England and Wales since 2003. Knee replacement data is collected on dedicated data collection forms for both 'first time' primary (K1 form) and 're-do' revision (K2 form) procedures (See Appendix ii). These forms have evolved over the lifespan of the NJR with information either added or withdrawn based on surgeon feedback. An overview of the information collected on the current versions of the K1 and K2 forms are summarised in table 3.3. Further information about the NJR can be found in sections 1.3 and 3.2.

For the current analysis the NJR provided data on all knee replacements performed between April 2003 and December 2010 for whom an NHS number was available. This database is equivalent to the NJR outlier database on which the NJR, via the University of Bristol, produce their outlier analysis for monitoring surgeon, centre and implant performance. Due to the requirement for an NHS number this database includes approximately 95% of all knee replacements registered with the NJR.

The database holds information on 416,628 primary knee replacements performed during the specified period. The information collected on the K1 form (table 3.3) was available for each record. Information on the status of the implant as of the 31st December 2010 was provided (implant unrevised, implant revised, and patient dead with implant *in situ*) in addition to the period for which it had been under observation. For those cases that had been revised the information collected on the K2 form (table 3.3) was also available. Details of the distribution of different types of knee replacement (total *versus* unicondylar *versus* patella-femoral *versus* 'other') and the associated numbers of revisions are given in figure 3.1.

	Available on K1 form	Available on K2 form
Patient Details		
- Patient consent	X	X
- Hospital ID	X	X
- Body Mass Index	X	X
Patient Identifiers		
- Name	X	X
- Gender	X	X
- Date of Birth	X	X
- Post Code	X	X
- NHS number	X	X
Operation Details		
- Hospital	X	X
- Operation Date	X	X
- Anaesthetic type	X	X
- American Society of Anaesthesiologists grade	X	X
- Operation Funding	X	X
Surgeon Details		
- Consultant in change	X	X
- Operating surgeon	X	X
- Operating surgeon grade	X	X
- First assistant grade	X	X
Procedure Details		
- Side	X	X
- Indication for surgery	X	X
- Pre-operative Range of Motion	X	-
Primary Operation Details		
- Primary date	N/A	X
- Primary Hospital	N/A	X
Components Removed		
- Brand of Knee removed	N/A	X
Surgical Details		
- Procedure type	X	X
- Surgical approach	X	X
- Minimally invasive surgery	X	-
- Computer assisted surgery	X	-
- Type of venous thromboprophlaxis	X	X
- Use of bone graft	X	X
Untoward intraoperative events	X	X

Table 3.3: Summary of the variables relating to knee replacement recorded by the NJR.

Key X = Recorded, - = Not recorded, N/A = Not applicable.

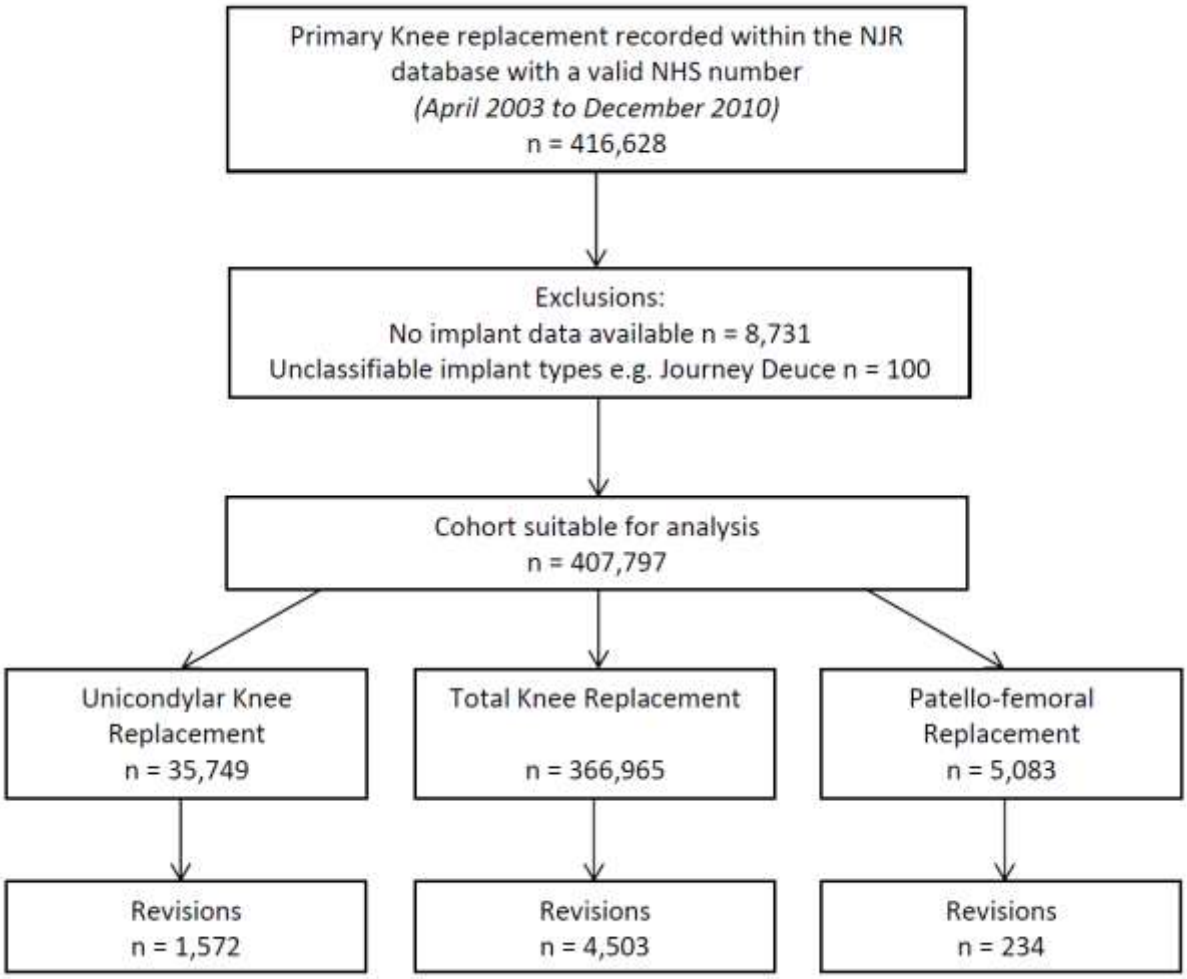


Figure 3.1: Distribution of different types of knee replacement and associated revisions within the NJR dataset. Note: Journey Deuce is a bicondylar implant comprising a combined unicondylar tibio-femoral and patella-femoral replacement.

3.4.3 Overview of the data available from the Patient reported outcome measures (PROMs) database

Traditionally knee replacements have been assessed using objective measures of the surgical and technical aspects of the procedure. These included implant survivorship using revision as an end-point, implant alignment and rates of complications. There are three main problems with this approach. Firstly these measures are determined by the surgical team based on clinical and radiological observations and are therefore prone to reporter bias. Secondly, they fail to take into account the patient's perspective of the results of surgery. Thirdly, revision as an endpoint tends to underestimate problems as patients can have a poor functional outcome without undergoing revision surgery (Noble 2006, Bourne 2010).

Unfortunately, a technically successful procedure does not guarantee a satisfied patient. The presence of continuing symptoms such as pain and functional limitation can lead patients to feel unhappy with the outcome of their surgery (Rothwell 1997, Janse 2004). The priorities and concerns of patients and surgeons after knee replacement often differ (Bullens 2001). Surgeons tend to be focussed on the relief of specific symptoms and return to defined activities whereas patients tend to consider more general improvements and the impact that their joint replacement has had on their overall quality of life (Allyson Jones 2005). An appreciation of this disparity has helped to shape the way in which we evaluate knee replacements with an increasing focus of patient reported outcomes within the assessment of surgical outcome (Garratt 2004, Jones 2005).

In 2008 Lord Darzi produced the High Quality Care for All: NHS Next Stage Review for the Department of Health (Darzi 2008). At the heart of the report was an aim to put quality at the heart of everything the NHS does. Particular importance was placed on quality as assessed by patients themselves. Patient Reported Outcomes Measures (PROMs) were

outlined as an integral part of the assessment of the effectiveness of care from the patient's perspective:

"*Effectiveness of care. This means understanding success rates from different treatments for different conditions. Assessing this will include clinical measures such as mortality or survival rates and measures of clinical improvement. Just as important is the effectiveness of care from the patient's own perspective which will be measured through patient-reported outcomes measures (PROMs)…*". (Darzi 2008)

PROMs measure the quality of the episode of care from the patient's perspective by recording aspects of functional improvement specific to the procedure being performed in addition to measurement of general health status and patient satisfaction (PROMs 2012). For knee replacement analysis these data complement existing registry based information and help to "*fill the gap in the set of information available on the care delivered to NHS-funded patients*" (Darzi 2008).

Potential uses of PROMs data include (Darzi 2008):

- Evaluation of the relative clinical quality of providers of elective procedures. PROMs can be used to benchmark performance.
- Informing patient choice, allowing patients to make informed decisions over their healthcare based on quality information.
- Producing research about what works. Efficacy and cost-effectiveness of different technical approaches to care can be evaluated using PROMs in association with other measures.
- Assessing the appropriateness of referrals to secondary care. PROMs data can be used to establish whether referrals for elective procedures are appropriate by examining variation in baseline PROMs scores across the country.

- Supporting the reduction of inequalities.
- Empowering commissioners.
- As a basis for financial remuneration through Payment by results (PbR) and the Commissioning for Quality and Innovation (CQUIN) payment framework.

In April 2009 the NJR, alongside the Department of Health (DoH), commissioned the collection of PROMs data to augment the information already held within the NJR. Since the programme started the DoH has been routinely collecting PROMS data before and after selected NHS funded elective surgery. Initially four procedures were chosen for PROMs collection namely total hip replacement, total knee replacement, varicose vein surgery and hernia repair, but work is under way to assess the feasibility of extending routine measurement to a range of chronic conditions. This project is intended to help the NHS measure and improve the quality of its care and will use patient's perspectives to inform decision making. This reflects key themes in current NHS reforms to put patients in charge of making decisions about their care (Darzi 2008). A limited amount of PROMs data for hip and knee replacement is currently published online at www.hesonline.org.uk. However, this is a crude analysis that combines all patients and all implant types together. For the orthopaedic community the opportunity exists to amalgamate patient outcome measures with clinical outcome data from the NJR, allowing a more sophisticated and comprehensive analysis of specific operations and subgroups of patients to be undertaken. This will supplement the information already available from the NJRs annual reports.

For joint replacement the PROMs questionnaires are administered pre-operatively (usually collected by the operating institution) and 6 months post-operatively (collected by postal questionnaire), and record information on general and condition specific measures of health status (Appendix iii). While collection is currently limited to these time points there are plans to extend collection to 1, 3 and 5 years for a subset of patients (NJR-AR 2012). General health status is measured using the Euroqol EQ5D which is common to all questionnaires permitting comparison both within and between procedures. The condition specific

measures relate to the procedure under investigation and hence can only be compared within that procedure. For knee replacement this is measured using the Oxford knee score (Dawson 1998). The generic and condition specific metrics were chosen after systematic review of all available measures of health utility and knee-specific outcome (Smith 2005).

The EQ5D was chosen as it is a short, concise and clear questionnaire that has been validated as a measure of health utility in a number of different scenarios and healthcare settings. In addition it has been endorsed by the National Institute for Health and Clinical Excellence which requires "that health states should be measured using a generic and validated classification system for which reliable UK population preference values, elicited using a choice-based method such as the time trade-off or standard gamble (but not rating scale), are available" (Smith 2005).

When considering the disease specific metric a range of scoring systems were considered and their reliability, validity and responsiveness assessed. This process is summarised in figure 3.2 taken from the report assessing which outcome measures were most pertinent for the PROMs project (Smith 2005). Psychometrically there was little to choose between the Western Ontario and McMaster Universities Arthritis Index (WOMAC) and the OKS. The WOMAC is widely used in the United States and has a high volume of associated literature, but is less widely used in the United Kingdom. In contrast the OKS was developed in the UK and is widely used in British orthopaedics. The OKS has acceptable reliability and responsiveness and most forms of validity. When questioned orthopaedic surgeons based in the United Kingdom expressed a preference for OKS on the basis of a perceived reluctance of patients to complete WOMAC which is longer and more time consuming to complete, increasing the possibility of missing data. On the basis of these findings the advisory panel therefore recommended the OKS as the knee specific metric for the PROMs project.

In addition to the EQ5D and Oxford score PROMs also collects other relevant information. As part of the pre-operative questionnaire patients are asked about co-morbidities, living arrangements and self-reported disability which can be used to understand the differences in health status between patients presenting for different surgical procedures. The post-operative questionnaire includes information on complications (UTI/Wound problems/Bleeding problems/Allergic reactions) (Audit Commission 1991), reoperations and readmissions as well as specific questions relating to satisfaction and the patients perception of how successful their surgery has been (Smith 2005). An overview of the data recorded as part of the PROMs programme is given in table 3.4. Details of the pre-operative variables available in the linked NJR-PROMs dataset are described in table 3.8. Further details of the post-operative outcome variables used for this work are given in section 3.5.

	WOMAC™	OHS	OKS	LEFS	PSI	AIMS	THOEQ	MCKNEE	HARRIS	HRS	MAYO	HAQ	VAI
Reliability: Internal consistency	+++	+++	+++	++	0	0	0	0	0	0	0	0	0
Reliability: Test-retest reliability	++	+++	+++	+++	+++	0	+	+	0	+++	0	0	0
Validity: Content validity	++	++	++	0	0	++	0	++	0	0	0	+	0
Validity: Criterion-related validity	0	0	0	0	0	0	0	0	+++	0	+++	0	0
Validity: Construct validity: Within scale analyses	++	+++	+++	++	0	0	++	0	0	0	0	0	0
Validity: Analyses against external criteria: Construct validity: Convergent/ Discriminant	+++	+++	++	++	++	+++	+	0	++	++	0	0	0
Validity: Construct validity: Known groups	+++	+++	+++	0	0	+	0	0	0	0	0	0	0
Validity: Other hypothesis testing	++	0	0	0	0	0	0	0	0	0	0	0	0
Responsiveness	+++	+++	+++	0	+++	++	0	++	0	+	0	+++	0
Interpretability	+	+	+	0	0	+	0	0	+	+	0	+	0
Acceptability	++	+	+	0	0	0	0	0	0	0	0	0	0
Feasibility/burden	++	0	0	+	++	0	0	0	0	0	0	0	+

Figure 3.2: Appraisal of psychometric (shaded) and operation (unshaded) criteria for disease specific PROMs for knee replacement (adapted from Smith 2005). The OKS (third column) was chosen as the disease specific metric of choice. Notes: For psychometric criteria 0 = not reported or no evidence in favour, + = limited evidence in favour, ++ = some acceptable evidence in favour, but some aspects fail criteria or not reported, +++ = acceptable evidence in favour. For operational criteria 0 = no evidence, + = some evidence.

Patient related explanatory variables	Time recorded
Demographic variables	
- Age	Both
- Gender (M/F)	Both
- Date of questionnaire completion	Both
Patient health variables	
- History of heart disease	Pre-operative
- History of hypertension	Pre-operative
- History of previous stroke	Pre-operative
- History of circulatory problems	Pre-operative
- History of lung disease	Pre-operative
- History of diabetes	Pre-operative
- History of kidney disease	Pre-operative
- History of nervous system disease	Pre-operative
- History of liver disease	Pre-operative
- History of cancer	Pre-operative
- History of depression	Pre-operative
- History of arthritis	Pre-operative
- General health rating	Both
- Self-reported disability	Both
- Living arrangements	Both
- Duration of symptoms	Pre-operative
Outcome variables (Further details given in section 3.4)	
- Euroqol-5D index	Both
- Euroqol-5D visual analogue scale	Both
- Oxford Knee Score	Both
- Patient satisfaction rating	Post-operatively
- Patient success rating	Post-operatively
- Operative event (allergy)	Post-operatively
- Operative event (UTI)	Post-operatively
- Operative event (bleeding)	Post-operatively
- Operative event (wound problem)	Post-operatively
- Operative event (reoperation)	Post-operatively
- Operative event (readmission)	Post-operatively
Other	
Number of times attended physiotherapy	Post-operative

Table 3.4: Summary of the variables recorded by the PROMs project which were available for the current analyses.

*Further information relating to the data held in the PROMs database can be accessed via the PROMs data dictionary (http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=289).

3.4.4 Preparation of the NJR-PROMs linked database

For the current work the Department of Health information centre provided information on all completed PROMs episodes up to February 2011. Upon receipt of the PROMs data four key identifiers common to both the NJR and PROMs databases were identified. These were the NJR index number, the NJR procedure identifier, the episode key and the year of entry of that operation onto the Hospital Episode Statistics database. These identifiers were extracted from the PROMs database and provided to the NJR data handler who was then able to extract the matching NJR records. The two datasets (PROMs and NJR) were then merged based on matching of these four variables common to both datasets. Adequacy of this merge was assessed by cross checking for discrepancies between these four variables in the merged file. Of a total of 51,558 records there were no discrepancies for any of the variables and therefore the merge was felt to be successful.

The data held within the linked NJR-PROMs database was a composite of the data outlined in tables 3.3 and 3.4. In total the merged NJR-PROMs extract included all 51,558 records. Included within this were all patients who had completed either part of the PROMs questionnaire (Pre or Post-operative) between 01/04/09 and 01/02/11.

Initial analysis of the NJR-PROMs data revealed a number of data issues. These fell into 4 categories:

1. Missing questionnaire data (Either pre or post-operative questionnaires were incomplete).
2. Missing information relating to the data of questionnaire completion meaning the time to follow up could not be determined.
3. Variation in the time of follow up for those records were this information was available.
4. Duplicate records.

Each of these issues was dealt with in turn to produce a clean, usable dataset.

Step 1: Removal of cases with missing questionnaires

Improvement in scores (OKS and EQ5D) is one of the key outcomes used by HES and local health authorities when making comparisons using PROMs. It was also the method of comparison used when the initial evaluation and feasibility of using PROMs to compare performance was undertaken (Browne 2007). Calculation of the improvement in score at an individual level requires knowledge of the pre-operative starting point and the final post-operative endpoint. It is therefore not possible to calculate the improvement in score if either the pre or post-operative questionnaires are missing. A decision was therefore taken to remove all cases with missing questionnaires, leaving only those with both the pre and post-operative PROMs questionnaires completed. Missing questionnaires were identified using a column within the PROMs dataset that records whether the questionnaires had been received by the Department of Health. Questionnaires are recorded as received if they are returned irrespective of whether all sections have been correctly or adequately completed. A received questionnaire may therefore still have some sections missing but will have data within it that makes it possible to compare at least one PROMs outcome for that individual.

In total this process removed 18,676 cases (Table 3.5). However, of these 10,633 were patients who had completed a pre-operative questionnaire but who had not yet reached the 6 month post-operative threshold. These patients were therefore not yet eligible to complete the post-operative questionnaire.

Questionnaire completion	Number	Percent
Neither complete	157	0.3%
One complete	18,519	35.9%
- Not yet reached 6 month post-operative threshold	10,633	20.6%
- Missing questionnaire	7886	15.3%
Both complete	32,882	63.8%
Total	51,558	100.0%

Table 3.5: Questionnaire completion rates for the 51,558 NJR-PROMs records.

Step 2: Removal of cases for which the date of questionnaire completion could not be ascertained

Information on the date of completion of each of the PROMs questionnaires is required if the timing of completion relative to the date of surgery is to be ascertained. This is important as if the pre-operative questionnaire is completed well in advance of surgery it is possible the clinical picture may have changed and the PROMs recorded may not be an adequate reflection of the patient's status at the time of surgery. Similarly post-operative questionnaires completed too early in the post-operative course may not detect improvements in PROMs as patients are still in the recovery phase whereas questionnaires completed much after much longer follow-up may not be comparable to improvements seen earlier in the post-operative course. It is important that periods of follow-up are equal to limit the confounding effects of time to follow-up within any comparative analyses. Cases for which the date of questionnaire completion could not be determined, were therefore excluded (n=2676) (Table 3.6).

Dates for Questionnaires	Number	Percent	Cumulative Percent
No date for either	71	0.2%	0.2%
Date for one questionnaire only	2605	7.9%	8.1%
Dates for both questionnaires	30,206	91.9%	100.0%
Total	32,882	100.0%	

Table 3.6: Presence of dates on the 32,882 records with both PROMs questionnaires

Step 3: Removal of cases for which the timing of questionnaire completion was out with a relevant range

To produce a homogeneous dataset it was decided to limit the time from the pre-operative questionnaire to surgery and the time from surgery to the post-operative questionnaire. This posed two questions:

1. Should the NJR or PROMs date of surgery be used to calculate these time periods?
2. What should these time periods be?

The date of surgery is recorded in both the NJR and PROMs databases. The method of collecting this data differs between databases. For the NJR database the date of surgery is recorded on the K1/K2 form completed at the time of surgery and is thus likely to be accurate. If this data is missing it is verified and added by data clerks at the time of entry into the NJR database. In contrast the PROMs date of surgery is recorded as part of the post-operative questionnaire and is recorded by the patients in response to the question "Please confirm when you knee operation took place?". This therefore relies on patients completing this section, returning the questionnaire and accurately recalling the date of operation. After review the date recorded in the NJR database was chosen as the reference for the date of operation based on the following observations:

1. The NJR dataset had a higher rate of complete surgical date data (NJR = 51,556 vs. PROMs = 30,221).
2. NJR data is less prone to reporting error as its completed on day of surgery by surgeon rather than 6 months after surgery by the patient.
3. All NJR dates for surgery were within a plausible range whereas a number of PROMs dates were not (e.g. some dates in the future).

Acceptable ranges for the periods of follow up were then defined. For the pre-operative PROMs an arbitrary 90 day cut-off for completion was chosen. This was chosen as it is similar to the time patients wait on current NHS waiting lists and should therefore represent the symptoms at the time a decision was made to list them for surgery. Questionnaires completed at more than 90 days prior to surgery were excluded as symptoms and corresponding levels of function may have changed during this period. The distribution of time of completion for the post-operative questionnaires is shown in Figure 3.3. Due to the design of the PROMs project the majority of post-operative PROMs were collected between 6 to 8 months. Functional improvements after TKR begin to plateau at 6 months following surgery but may continue to improve for up to 12 months (Nerhus 2010). While comparison of outcomes after knee replacement are probably best compared at a minimum of 12 months after surgery it was recognised that unless patient with at least 6 months follow up were included then a significant volume of follow up data would be lost. Additionally there

is evidence that the functional improvements attained between 6 and 12 months are minimal and that comparison of outcome at 6 months is therefore acceptable (Judge 2011).

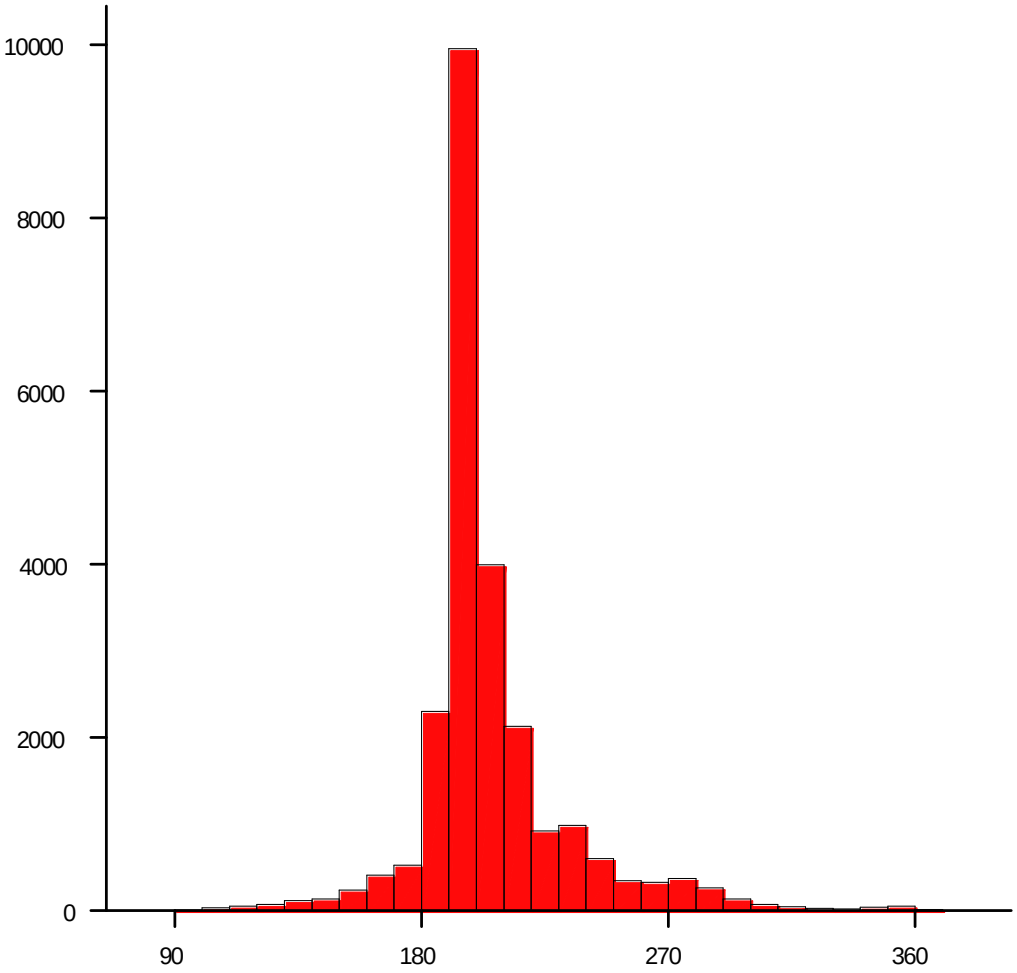


Figure 3.3: Distribution of the time in days of collection for the 30,206 post-operative PROMS records with both questionnaire complete and complete information for the dates of questionnaire completion.

Based on these criteria the dataset was filtered by the timing of pre- and post-operative questionnaire completion (Figure 3.4). In total 25,011 of the remaining 30,206 completed their pre-operative questionnaire within 90 days of surgery and their post-operative questionnaire within 6 to 12 months of surgery.

Step 4: Removal of duplicate records

Duplicate data were assessed on the basis of duplicate NJR ID numbers. The number of duplicates was small (7 of 25011 records (0.002%)). Where duplicates were found the second of the two duplicate records were removed from the dataset. The process for cleaning the data is summarised in figure 3.4. The number of duplicates and the final numbers of useable records for each implant type are given in figure 3.5.

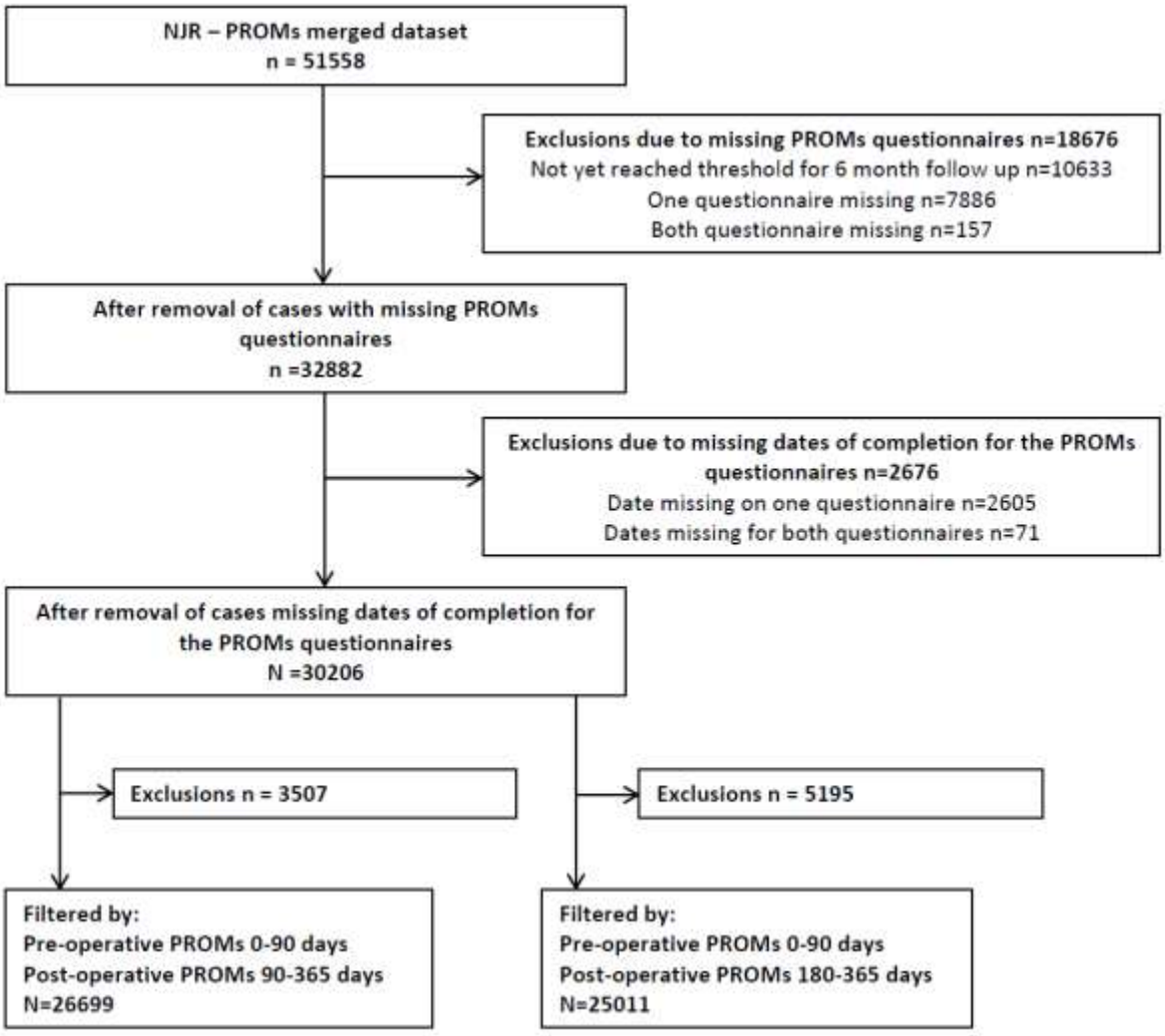


Figure 3.4: Flowchart summarising the preparation of the linked NJR-PROMs dataset.

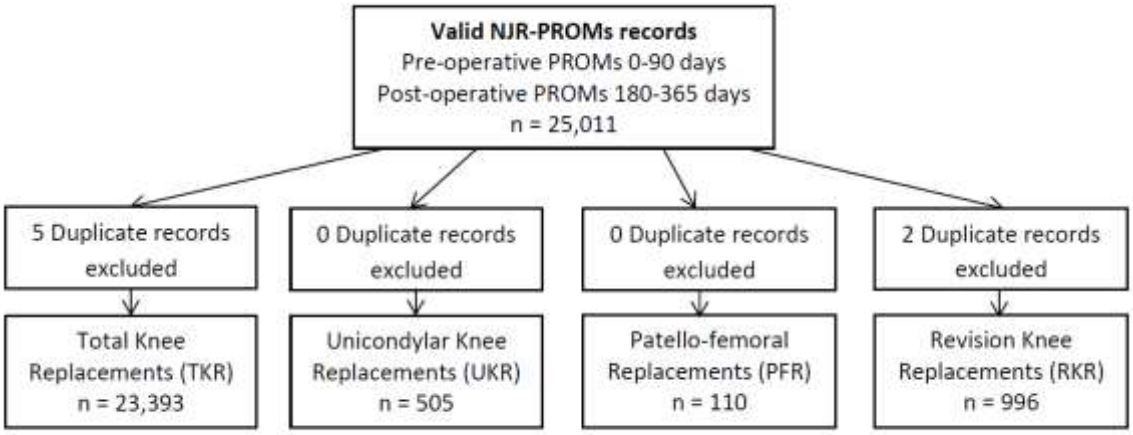


Figure 3.5: Details of the 7 duplicate records from the cleaned NJR-PROMs linked dataset.

Following removal of the excluded cases the demographics of the patients included and excluded were compared to ensure the patients with complete questionnaires were demographically comparable to the total PROMs population. The two groups were matched for age (p=0.54), gender (p=0.93) and side of operation (p=0.64). However, there were differences between the groups for the American Society of Anesthesiologist (ASA) grade (p<0.001) with a greater proportion of ASA 1 and 2 patients and a smaller proportion of ASA 3 patients observed in the group that was used for analysis (ASA grade 1: 10.5% *versus* 9.1%, ASA grade 2: 74.3% *versus* 73.1%, ASA grade 3: 14.8% *versus* 17.3%). A possible explanation for this may be that patients with significant co-morbidities (ASA 3 or more) had their surgery postponed due to these co-morbidities. If this occurred after the pre-operative questionnaire was completed then by the time surgery was rescheduled they may have been outside the 90 day limit for pre-operative questionnaire completion and they may therefore have been excluded.

To ensure that the data we were analysing represented the wider population undergoing primary knee replacement, the TKRs (n=23,393) and UKRs (n=505) from the selected NJR-PROMs cohort was compared with the demographic details for national TKR and UKR

patients recorded in the NJR 9th Annual Report (table 3.7). This demonstrated that the cohorts were comparable for both groups, suggesting that the patients selected were representative of the wider population of patients undergoing these two operations.

Preliminary analysis of the cleaned NJR data revealed that for some of the factorial data fields there was one overwhelming level and a number of others for which the numbers were relatively small. Prior to any analysis we therefore we chose to combined data within these groups. Consequently the grade of lead surgeon and indication for surgery were grouped into their most frequently occurring levels of Consultant and osteoarthritis, respectively, *versus* all other levels combined. This approach also meant that the factors could each be treated as covariates during analysis. The demographic and pre-operative data available from the NJR-PROMs dataset and a description of this data is given in table 3.8.

	Selected PROMs cohort		Comparison PROMs cohort TKR vs. UKR	Comparative values for 2010 from NJR Annual Report	
Variables*	**TKR (n = 23 393)**	**UKR (n = 505)**	**p-value**	**All TKR (n = 69 649)**	**All UKR (n = 6119)**
Mean (SD) age (yrs)	69.6 (9.0)	63.6 (9.8)	< 0.001	70.1 (9.3)	64.0 (9.8)
Mean (SD) Body Mass Index (kg/m^2)	31.0 (5.5)	30.1 (5.2)	0.005	30.7 (5.5)	30.0 (5.1)
Body Mass Index category (n, %)					
- 15 to 25 kg/m^2	1365 (*6*)	39 (*8*)			
- 25 to 40 kg/m^2	11 600 (*50*)	284 (*56*)			
- 40 to 60 kg/m^2	1036 (*4*)	17 (*3*)			
- Data missing	9392 (*40*)	165 (*33*)			
Gender (n, %)					
- Female	13 221 (*57*)	258 (*49*)	0.001	37 997 (55)	2675 (44)
- Male	10 172 (*43*)	247 (*51*)		28318 (41)	3027 (49)
- Missing				3334 (5)	417 (7)
ASA grade (n, %)					
- 1	2362 (*10*)	112 (*22*)	< 0.001	7455 (*11*)	1421 (*23*)
- 2	17 445 (*75*)	355 (*70*)		51 158 (*73*)	4221 (*69*)
- 3 or 4	3586 (*15*)	38 (*8*)		11 036 (*16*)	477 (*8*)
Side (n, %)					
- Left	11 186 (*48*)	242 (*48*)	0.96		
- Right	12 207 (*52*)	263 (*52*)			
Hospital type (n, %)					
- NHS	20 932 (*89*)	379 (*75*)	< 0.001	48 575 (*70*)	3645 (*60*)
- Independent (private/ISTC)	2461 (*11*)	126 (*25*)		21 074 (*30*)	2474 (*40*)
Lead surgeon grade (n, %)					
- Consultant	17 374 (*74*)	434 (*86*)	< 0.001		
- Other	6019 (*26*)	71 (*14*)			
Comorbidities† (n, %)					
- 0 or 1	17 771 (*76*)	428 (*85*)	< 0.001		
- ≥ 2	5622 (*24*)	77 (*15*)			
General health pre-operatively (n, %)					
- Excellent	818 (*4*)	22 (*4*)	0.06		
- Very good	5834 (*25*)	150 (*30*)			
- Good	10 440 (*45*)	222 (*44*)			
- Fair	4983 (*21*)	85 (*17*)			
- Poor	745 (*3*)	14 (*3*)			
- Data missing	573 (*2*)	12 (*2*)			
Self-assessed as 'disabled' (n, %)	12 925 (*55*)	175 (*35*)	< 0.001		
Diagnosis of depression (n, %)	1703 (*7*)	44 (*9*)	0.23		
Osteoarthritis as indication (n, %)	22 798 (*97*)	499 (*99*)	0.06	67 668 (*97*)	6048 (*99*)
Symptom period (n, %)					
- < 1 year	1166 (*5*)	40 (*8*)	< 0.001		
- 1 to 5 years	12 151 (*52*)	307 (*61*)			
- 6 to 10 years	5020 (*22*)	87 (*17*)			
- > 10 years	4969 (*21*)	71 (*14*)			
- Data missing	87 (*0*)	0 (*0*)			

Table 3.7: Comparison of the TKR and UKR cohorts form the final NJR-PROMs dataset with the equivalent patients undergoing these procedures described in the NJR 9th annual report. p values for the difference in demographics between the selected TKR and UKR patients are given to highlight the inherent difference between these two groups of patients. *Further details for each of the variables can be found in table 3.8.

Variable	Data source	Data type	Additional Information
PATIENT VARIABLES			
Age (years)	NJR/PROMs	Continuous/ categorical	Used either as a continuous variable or categorised into age groups (<55/55-64.9/65-74.9/≥75 years) determined by the spread of the data
BMI (kg/m2)	NJR	Continuous	Only BMIs within range 15-60 included
Number of comorbidities	PROMs	Continuous	Recorded by patients as part of the pre-operative PROMs questionnaire. Total of 10 comorbidities: heart disease, hypertension, stroke, circulatory problem, lung disease, diabetes, kidney disease, nervous system disease, liver disease, cancer History of depression and arthritis excluded as history of depression considered separately (see below) and history of arthritis overlapped with indication for surgery.
Time from operation to post-operative PROMs questionnaire collection (days)	NJR/PROMs	Continuous	Calculated from date of operation as recorded on the NJR database to date of post-operative PROMs as recorded on the PROMs questionnaire
Gender	PROMs	Binary	Male/Female
Side of surgery	NJR	Binary	Right/Left
Pre-operative disability	PROMs	Binary	Indicates whether the patients considers themselves to have a disability (Yes/No)
Pre-operative general health	PROMs	Ordinal	Indicates the patients perception of their own general health with 5 ordered options: 1. Excellent 2. Very good 3. Good 4. Fair 5. Poor
Depression	PROMs	Binary	Indicates whether patients have previously been given a diagnosis of depression
Anxiety level	PROMs	Ordinal	Derived from the Anxiety/depression component of the EQ5D index. Indicates their current level of anxiety/depression with 3 ordered options: 1. I am not anxious or depressed 2. I am moderately anxious or depressed 3. I am extremely anxious or depressed
ASA grade	NJR	Ordinal	American Society of Anesthesiologist grade 1 to 5
Indication for surgery	NJR	Binary	Osteoarthritis vs. Other indication – Indication for surgery included >10 different indications. Decision taken to re-code this data into a binary grouping based on the observations that a) osteoarthritis was the indication in >95% of all operations b) all of the other indication accounted for <1% when considered individually
Duration of symptoms	PROMs	Ordinal	Indicates the duration of knee symptoms with 4 ordered options: 1. <1 year 2. 1-5 years 3. 6-10 years 4. >10 years
SURGICAL VARIABLES			
Lead surgeon grade	NJR	Binary	Consultant vs. Other grade – Lead surgeon information included 10 different surgical grades. Decision taken to

			re-code this data into a binary grouping based on the observations that a) consultants performed >70% of all operations and the numbers for each of the other groups were therefore small b) interest lies in knowing whether not having a consultant perform you operation influences outcome, irrespective of the grade of the operating surgeon if they are not a consultant
Hospital type	NJR	Nominal	NHS funded patients either operated in • NHS hospital • Independent hospital • Independent Surgical Treatment Centre (ISTC)
Type of Knee replacement	NJR	Nominal	4 broad groups identified • Total Knee Replacement (TKR) • Unicondylar Knee Replacement (UKR) • Patello-femoral Replacement (PFR) • Revision Knee Replacement (RKR)
Bearing	NJR	Binary	Fixed/Mobile bearing – Information obtained from the component codes held within the NJR
Meniscus (For TKR only)	NJR	Nominal	Information obtained from the component codes held within the NJR. Information grouped into three groups: • Cruciate Retaining • Posterior Sacrificed • Other
Patella Resurfaced (For TKR only)	NJR	Binary	Indicates whether the patella was resurfaced (Yes/No)
Brand type (For TKR only)	NJR	Nominal	Information obtained from the component details held within the NJR. The top 5 brands were chosen for analysis as they constituted 80% of all TKR alongside an 'other' group: • PFC® (Depuy) • NexGen® (Zimmer) • Genesis 2® (Smith and Nephew) • AGC® (Biomet) • Triathlon® (Stryker) • Others
Minimally invasive surgery	NJR	Binary	Yes/No
Computer Navigated	NJR	Binary	Yes/No
PRE-OPERATIVE SCORE DATA			
Pre-operative OKS	PROMs	Continuous	0 worst to 48 best
Pre-operative EQ5D	PROMs	Continuous	Max score 1, Scores <0 indicate a state worse than death
Pre-operative VAS	PROMs	Continuous	Visual analogue scale scored 0 (worst) to 100 (best)

Table 3.8: Summary of the explanatory patient, surgical and score variables available in the NJR-PROMs linked dataset.

3.5 Description of the outcome variables used in the current analysis

Linking the NJR-PROMs datasets meant that a number of different outcome variables became available and could be used for as the basis for the current analyses. The variables available are described below and include:

1. Implant revision.
2. Reason for revision.
3. Functional outcome scores.
    a. Oxford Knee Score.
    b. Euroqol 5D.
4. Patient reported satisfaction / success.
5. Rates of post-operative complications.

3.5.1 Implant Revision

Implant revision is the cornerstone of registry analysis. The NJR defines revision as *"exchange of one implant for another or removal of implants as part of a staged procedure*". The life of every knee replacement recorded with the NJR is monitored from the time of implantation until the implant fails or the patient dies. At any particular point in time the period for which an implant has been under observation and its status (implant unrevised, implant revised and patient dead with implant *in situ*) can be determined. For an implant to be coded as revised a revision procedure has to be recorded on the NJR database and linked to the relevant primary procedure using unique NJR identifiers (NJR-NJR linkage). In the early years of the registry 2003 to 2009 an alternative way of identifying revision was employed whereby revisions were identified in the Hospital Episode Statistics (HES) database (NJR-HES linkage).

The influence of the different linkage methods on the revision rate was described in the 8th NJR Annual Report. This concluded that "*while NJR-NJR linkage underestimates revisions to some undetermined extent it was likely that HES data is likely to overestimate revisions to*

some degree because of the inclusion of some re-operations as revisions. It is likely then that the "real" revision rate lies somewhere between the two rates". Therefore it seems that neither method is any better than the other as both methods have inherent flaws and, because there is no 'perfect' way of measuring revision rate it is impossible to quantify if one of these methods is any better of worse than the other. Presently, NJR-NJR linkage is the method used by the NJR for their outlier analysis of surgeons and centre survival and will be the only way revisions are reported from the 9th NJR annual report onwards. It is also the method used in a number of recent research papers published using NJR data (Jameson 2012, Smith 2012). In line with current NJR policy we have therefore used NJR-NJR linkage to identify revision procedures within the analyses presented in this thesis.

3.5.2 Reason for revision

For those knee replacements that undergo revision it is important to understand why the knee was revised. Differences in the mechanism of failure between implants might help to explain why one implant fails to a greater or lesser extent than another. It is also important to recognise as the mode of failure may influence survival and function of the revised implant. Information relating to the reason for revision is recorded on the K2 and is therefore available in the NJR database. The K2 form is completed by the surgical team immediately following revision surgery. The reason for revision is therefore recorded based on pre-operative clinical and radiological investigations alongside per-operative surgical findings. This allows the surgeon ample opportunity to determine the exact mode of failure. It also allows surgeons to choose more than one option when recording the reason for revision if multiple reasons become apparent at the time of surgery.

To overcome the problem of multiple reasons for revision we employed a hierarchical strategy for determining the primary reason for revision. This mirrored the hierarchy used by the Australian Arthroplasty register (AJR-AR 2009) but was modified to accommodate the additional reasons for revision available on the NJRs data collection forms (table 3.7). In the situation where multiple reasons were stated the revision was attributed to the highest

ranking reason. As an example, it can be seen from table 3.9 that unexplained pain is the lowest ranked of the predefined reasons and as such only revisions where unexplained pain was stated alone were found in this category. Where unexplained pain was seen in combination with an additional higher ranked reason the revision was attributed to this additional reason.

Hierarchy	Reason for Revision	Category
1	Tumour	Dominant diagnosis independent of prosthesis
2	Infection	Dominant diagnosis independent of prosthesis
3	Incorrect side or sizing of component	Related to surgical procedure
4	Malalignment	Related to surgical procedure
5	Metal sensitivity	Reaction to prosthesis
6	Loosening / Lysis	Reaction to prosthesis
7	Component dissociation	Wear and implant breakage
8	Component wear / Polyethylene wear	Wear and implant breakage
9	Implant breakage	Wear and implant breakage
10	Dislocation / Instability	Stability of prosthesis/ knee
11	Fracture of bone	Fracture of bone
12	Progression of arthritis/disease	Progression of disease in non-operated part of joint
13	Synovitis	New disease occurring in association with prosthesis
14	Arthofibrosis / Stiffness	New disease occurring in association with prosthesis
15	Osteonecrosis/AVN	New disease occurring in association with prosthesis
16	Heterotopic bone	New disease occurring in association with prosthesis
17	Unexplained pain	Pain
18	Other (if not listed above)	Remaining diagnoses
19	No reason stated	

Table 3.9: Hierarchy of revision used to determine primary reason for revision where more than one reason for revision was recorded (adapted from Australian National Register (AJR-AR 2009)).

3.5.3 Functional outcome scores

A. Oxford Knee Score

The Oxford Knee Score (OKS) is a disease specific patient administered questionnaire exploring the patient's subjective assessment of pain, function and ability to perform activities of daily living (ADL's) following knee replacement surgery (Dawson 1998). It is a 12-part questionnaire (table 3.10) assessing patient perspectives of those factors linked to outcome. The answer to each question is rated on a scale ranging from 0 to 4 with lower scores indicating more severe problems. The scores for each question are added to generate an overall score between 0 and 48, with 0 representing the worst possible score, and 48 the best possible score.

In a review of health instruments for the assessment of the knee the OKS was found to be a reliable (extent to which the items of the instrument measure the entity, stability of the instrument over time), valid (extent to which the instrument measures what it is supposed to measure) and responsive (ability of instrument to measure significant change over time) measurement when compared to other regularly used patient reported outcome measures (Garratt 2004). Twinned with the fact that it is simple, easy to administer and suitable for following up patients in the longer term it has been recommended as an appropriate disease specific tool for assessing outcomes following total knee replacement (Davies 2002). It has also been recommended in preference to the WOMAC for large knee arthroplasty databases in a cross sectional population (Dunbar 2001). The method by which the OKS was chosen as the disease specific metric for the PROMs project is discussed in section 3.4.3.

Possible weakness with the OKS include the influence of coexistent hip or spinal pathology, which can significantly alter both the absolute score and any improvement to be expected after surgery (Harcourt 2001). Some patients have also identified areas of weakness within the score including a lack of question clarity, difficulty in reporting measurements of pain and restrictive and irrelevant questions. They also highlighted the influence and effects of co-morbidity on response (Wylde 2005).

Question number	Question
1	Describe the pain you usually have from your knee?
2	How much trouble do you have washing and drying yourself?
3	How much trouble do you have getting in/out car or using public transport?
4	How long can you walk before pain becomes severe?
5	After meal how painful has it been to stand up from a chair?
6	Have you been limping when walking?
7	Could you kneel down and get up again?
8	Have you been troubled by pain in bed at night?
9	How much has pain from your knee interfered with your normal work?
10	Have you felt your knee might suddenly give way or let you down?
11	Could you do the shopping on your own?
12	Could you walk down a flight of stairs?

Table 3.10: Summary of the component questions of the Oxford Knee Score (Also see Appendix iii).

B. Euroqol-5D (EQ5D)

The EQ5D is a simple, generic measure of health used for clinical and economic appraisal. It provides two separate measures of general health, the EQ5D index and the EQ5D VAS. The EQ5D index assesses five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) each graded 1 to 3 (Table 3.11 and Appendix III). Level 1 represents no problems and Level 2 and 3 represents difficulties, either moderate (Level 2) or severe (Level 3). Outcomes of this score can be reported in a table of the proportion of patient in each category before and after the intervention under investigation. In this method of reporting the final scores do not have arithmetic properties and are therefore not suitable for use as a cardinal score. Alternatively the scores can be combined using population weighting to produce the EQ5D index, a single measure of health status with a maximum possible value of 1 and a score below 0 indicating a state worse than death. The EQ5D health VAS (0-100 Visual Analogue Scale) provides an additional assessment of patient well-being.

EQ5D domain	Possible responses
By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.	
Mobility	- I have no problems in walking about - I have some problems in walking about - I am confined to bed
Self-Care	- I have no problems with self-care - I have some problems washing or dressing myself - I am unable to wash or dress myself
Usual Activities (e.g. work, study, housework, family or leisure activities)	- I have no problems with performing my usual activities - I have some problems with performing my usual activities - I am unable to perform my usual activities
Pain/Discomfort	- I have no pain or discomfort - I have moderate pain or discomfort - I have extreme pain or discomfort
Anxiety/Depression	- I am not anxious or depressed - I am moderately anxious or depressed - I am severely anxious or depressed

Table 3.11: Component parts of the EQ5D index (Also see Appendix iii).

3.5.4 Patient reported Satisfaction / Success

The PROMs questionnaire asks about patient satisfaction with their surgery by asking: "How would you describe the results of your operation?". Responses are recorded on a 5 point Likert scale with the possible answers: 'Excellent', 'Very Good', 'Good', 'Fair' and 'Poor'. Success is assessed by asking patients "Overall, how are the problems now in the knee on which you had surgery, compared to before your operation?" with possible responses 'Much better', 'A little better', 'About the same', 'A little worse', 'Much worse'. Success measures the patient's perception of whether they have symptomatically improved following surgery whereas satisfaction measures the extent to which they are happy with this improvement.

In both of these scales the first category is the best response and the fifth category the worst response. The ordinal scales used to assess satisfaction and success have not been validated. However, they do mirror similar adjectival scales used for assessing patient reported satisfaction in national cohorts (Robertsson 2000, Noble 2006, Howie 2010). The benefit of these scales is that they give a simple representation of the patient's perceptions of the results of surgery. In this respect they complement some of the more frequently used

validated outcome scores by appreciating the individual patient's experiences, instead of focusing on hard symptomatic endpoints.

3.5.5 Rates of post-operative complications

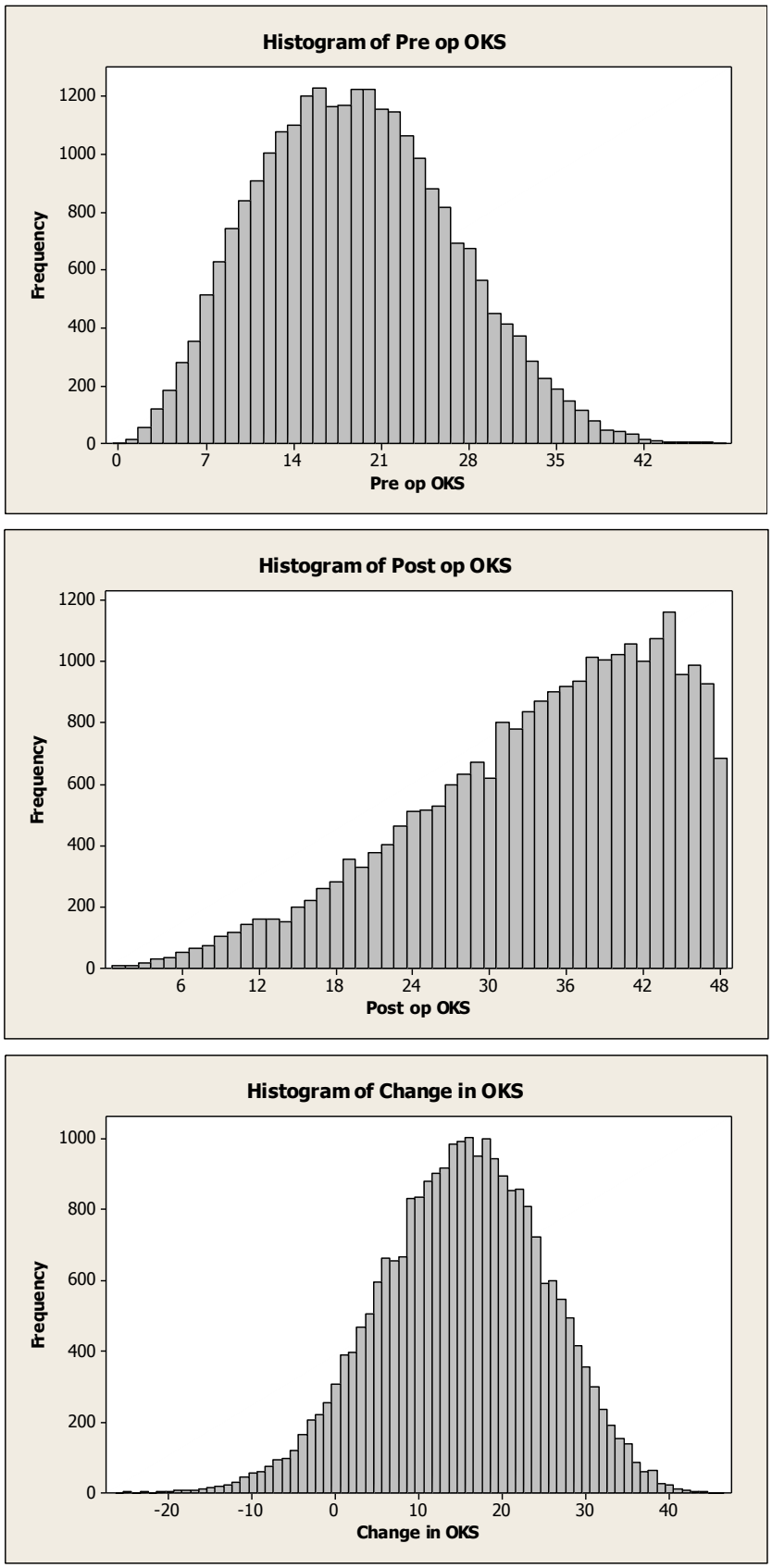
As part of their post-operative assessment patients are asked if they developed any of 6 specific complications following surgery (allergic reaction, urinary problems, wound problems, bleeding problems, readmission and need for further surgery). These questions form part of the patient's experience of surgery questionnaire which has previously been used to audit complications after day case surgery (Audit Commission 1991, Smith 2005). The interpretation of what constitutes a complication is at the discretion of the patient and no attempt is made to verify the presence of these complications from the medical records. This information is particularly useful when comparing groups for which the interpretation of a complication is expected to be similar such as after different type of knee replacement.

3.6 Statistical considerations

To achieve the objectives of each analysis a number of complementary statistical techniques were employed. An overview of these techniques is given in the following section. The details of the statistical methods used for specific analyses are discussed within the chapters relating to those analyses (Chapters 4 to 6).

3.6.1 Basic data analysis and simple comparative analysis

Prior to analysis the data was interrogated using graphical (scatterplots, histograms, interval plots etc.) and tabular summaries along with simple summary statistics (mean, median, standard deviation, range). This determined the distributions and characteristics of these data so that appropriate statistical analysis could be employed. For example, analysis of the Oxford Knee Score data demonstrated that the pre-operative score was approximately normally distributed with a slight right skew, the post-operative OKS was heavily skewed to the left with an abrupt cut-off at the top end of the scale (score 48), and that the change in score was approximately normally distributed (figures 3.6 to 3.8). A similar effect was seen for both the EQ5D index and EQ5D VAS. These findings demonstrated that the change in score was a much better outcome variable to model that the change in score based on its superior distribution characteristics. For comparative analysis the continuous variables were analysed using independent Student's t-tests, and one-way analysis of variance (ANOVA). Within ANOVA post hoc between groups comparisons were performed using the Bonferroni method. Categorical data was analysed using Fisher's exact test and Chi squared tests.



Figures 3.6, 3.7 and 3.8: Histograms deonstrating the distibutions of the pre-operative (top), post-operative (middle) and change (bottom) in Oxford Knee Score (OKS).

3.6.2 Survival analysis

For all of the survival analyses a revision procedure, as defined and recorded by the NJR was considered to be a 'failure event'. For each implant its status (Revised/Unrevised/Patient death with implant *in situ*) and the period for which it had been under observation was determined based on a census date of 31st December 2010. The time between the index procedure and any subsequent revision measured the joint's survival. Information on patients who had died before the census date was obtained through the United Kingdom Office for National Statistics (ONS).

Survival analyses were conducted in accordance with the recommendations of the Nordic Arthroplasty Register Association (NARA) study group (Ramstam 2011). Survival analyses are often conducted using one of two approaches, the life-table approach (Cutler 1958) and the Kaplan-Meier (product-limit) approach (Kaplan 1958). In both methods the survival is calculated as the cumulative probability that an implant will survive through a set of time points. However, they differ in that life-tables are calculated for a set of pre-defined time intervals whereas Kaplan-Meier analysis defines the time intervals to include only one event (revision) and is therefore independent of a subjective choice of time intervals (Ramstam 2011). While Kaplan-Meier analysis is usually preferred (Ramstam 2011) both methods have been advocated for the reporting of arthroplasty data (Murray 1993). Both Kaplan-Meier and life-table analysis methods were employed in the reporting of our survival data (Armitage 1994).

One issue with survival analysis is the influence of the competing risk of death. This problem particularly effects survival analysis performed over longer and longer periods as more and more patients are likely to die with time. It has been argued that the presence of a risk of a competing event may bias Kaplan-Meier survival estimates (Biau 2007). However, others have argued that because standard Kaplan-Meier analysis is based on the assumption that patient will be alive until the implant fails it give a more logical, understandable and clinically relevant survival estimate (Ramstam 2011). Currently the NARA group recommend that it is

appropriate to use Kaplan-Meier methods, however if this method is used the number and type of censored observations should be described (Ramstam 2011). For these analyses we have therefore used the standard Kaplan-Meier methods in addition to reporting why observations were censored (revision or death).

For a number of analyses we were interested in comparing survival between a large numbers of groups (up to 900 groups). In this situation conventional Kaplan-Meier plots and life tables are impractical. This problem was overcome by calculating the number of revisions per 100 component years for each group. This was done by firstly determining the cumulative period of observation for all implants within a given group. Knowledge of the number of revisions within this group during the period of observation is then used to calculate the number of revisions per 100 component years. This information could then be graphically represented using funnel plots allowing a visual comparison of the data to be made (Spiegelhalter 2005). For the purposes of these analyses funnel plots were constructed using +/- 3 standard errors of the mean which can be considered as 99% confidence intervals for the mean.

For a number of analyses we were interested in whether there was a difference in survival between one or more groups. Survival comparisons were made using the Mantel Hantszel Log-rank test (Mantel 1959). The benefit of this test is that it compares survival based on the total follow-up rather than at a pre-specified point chosen by the investigator. The choice of time point can influence the estimated difference in survival between groups making it a possible source of bias and by using the log-rank test this problem is mitigated.

For many comparisons there may be systematic differences in demographics of the groups being compared. Patient and surgical factors may influence implant survival. Differences in these variables between groups may therefore affect the validity of results by confounding bias (Ramstam 2011). This problem was partly overcome using Cox proportional hazard models. In these models, the contribution of potential risk factors to the risk of the event

occurring can be quantified (Therneau 2000) by determining the extent to which the timing of event is explained by the measured patient and surgical factors. The Cox model assumes that there is an underlying unspecified baseline hazard that stays constant through time and that is influenced by covariates that mitigate or enhance the risk of failure. The model is based on the assumption that the hazards of failure for comparative groups are proportional. Results obtained using this method are biased if the proportionality assumption is violated (Ramstam 2011). The main reason for non-proportionality is that a factor's effect on survival varies with time. Prior to modelling we therefore investigated the constant proportionality over time assumption by assessing the influence of time on each of the factors included within the models. This was achieved by creating time-dependent risk factors which were included in the Cox models.

Cox-models were used in two specific ways. Firstly, they were used to investigate which factors influence the hazard of failure. This was achieved using forward (from the null model) and backward (from the full model) stepwise regression. Having determined the factors that influenced survival the model was re-evaluated as a directly entered model (non-stepwise) to provide unconditional factor estimates. Secondly, they were used to adjust for differences in patient and surgical factors that could potentially confound comparison of a pre-specified group. In this situation a directly entered model (non-stepwise) was created including all of the 'adjuster' variables with the comparative group of interest added last to provide factor estimate for the group of interest conditional on the 'adjuster' variables. In all of the final models the combined influence of variables was determined by exploring 2-way interactions between the included covariates.

3.6.3 Generalised linear modelling

In similarity to survival comparisons, the validity of PROMs comparisons between groups may also be confounded if there are systematic differences in demographics of the groups being compared. Generalised linear modelling was therefore used to explore the influences of a number of explanatory variables upon our key PROMs outcomes, namely the change in

Oxford Knee Score, the change in EQ5D index, the presence of post-operative complications, the rates of patient satisfaction. Once the factors influencing the PROMs and the size of their effects were known the models could then be used to make adjustments for any differences in these factors between the comparative groups of interest.

Generalised linear modelling is a term relating to the unified approach to the analysis of a whole host of useful regression models including the normal linear regression model, the binary and Poisson regression models and a range of others models (OU 2009). For data that was approximately normally distributed (OKS change, EQ5D change) multivariable linear regression models were used. For data that was binary or ordinal (complication rate, satisfaction) ordinal logistic regression models we used.

For all modelling a standardised procedure was employed:

Step 1: Graphical analysis of the explanatory and outcome variables to characterise their distributions. This enabled appropriate models for the outcomes of interest to be chosen and allowed explanatory variables to be entered correctly into the model (either and continuous, ordinal, nominal or binary data).

Step 2: Exploration of the relationships between variables to be included in the model using scatterplot matrices and correlation matrices (figure 3.9 and table 3.12). This gives an overall impression of the strength of relationship between variables and therefore which variables are likely to appear in the final regression models.

Step 3: The influence of each explanatory variable in isolation was explored by creating univariate regression models.

Step 4: The influence of variables in combination was explored by creating ‘full’ multivariate models.

Step 5: Models were then limited to only the variables with a significant influence using backward stepwise regression from the ‘full’ model. The modelling process was repeated

using forward stepwise regression from the 'null' model to assess the effect of the direction of modelling on the variables included in the final model. Where there was discrepancy variables were added and removed from the model individually and their influence assessed subjectively based on the amount of variation in the outcome variable they explained (Wald test statistic) and their significance level (p-value). The significance levels employed varied dependent upon the type of analysis and the size of the dataset. p-values of $p<0.01$ to $p<0.001$ were used given the large number of patients and variables available to ensure a parsimonious model was achieved and to limit the effects of type I error.

Step 6: Having determined the variables that had a significant influence the models were then re-evaluated as a directly entered model (non-stepwise) to provide unconditional estimates for each variable. These models were then used to make adjusted comparisons between groups of interest.

Step 7: The influence of interactions between variables within the final models were assessed by investigating their 2 way interactions. Significant interactions were investigated graphically to quantify their effects.

Step 8: Linear regression model adequacy was assessed using standard analysis of residuals, leverage and Cooks statistics (figure 3.10). Ordinal logistic regression model identifiability was assessed on the basis of a condition number of Hessian of $<10^4$. Where the condition number of the Hessian was in excess of this value the models were rejected on the basis of being ill-conditioned. Ordinal logistic regression assumes that the relationship between each pair of outcome groups is proportional. This means that the impact of any covariate on each step up (or down) the ordinal scale remains the same with equivalent coefficients (and hence parallel regression lines across all levels of the ordinal scale). The extent to which the proportionality assumption was valid was assessed using the graphical methods described by Bender (1997) and Gould (2000) (figure 3.11).

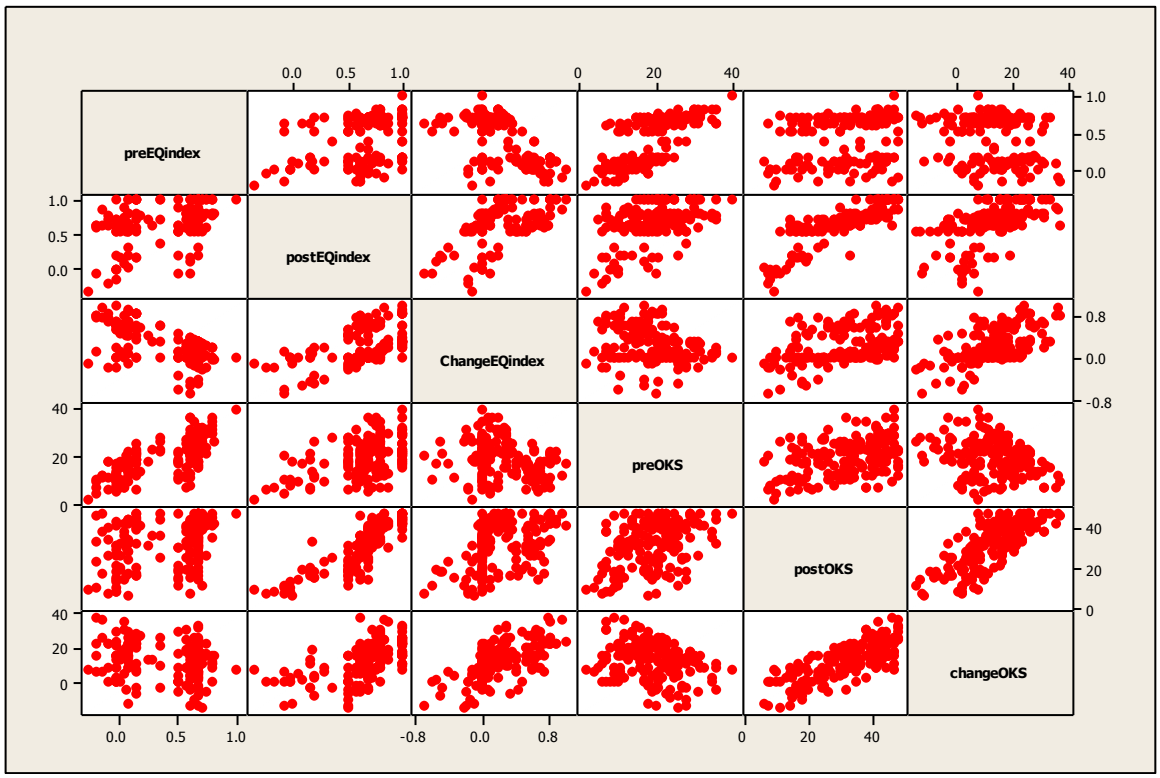


Figure 3.9: Scatterplot matrix demonstrating the relationships between the pre, post and change in EQ index and the pre, post and change in the OKS

	PreEQindex	PostEQindex	ChangeEQindex	PreOKS	PostOKS	ChangeOKS
PreEQindex	-					
PostEQindex	0.359 p<0.001	-				
ChangeEQindex	-0.612 p<0.001	0.518 p<0.001	-			
PreOKS	0.696 p<0.001	0.369 p<0.001	-0.326 p<0.001	-		
PostOKS	0.344 p<0.001	0.781 p<0.001	0.347 p<0.001	0.408 p<0.001	-	
ChangeOKS	-0.137 p<0.001	0.556 p<0.001	0.597 p<0.001	-0.280 p<0.001	0.762 p<0.001	-

Table 3.12: Correlation matrix demonstrating the correlations between the pre, post and change in EQ index and the pre, post and change in the OKS. Values in each box are the Pearson correlation co-efficient between the two variables and the associated p-value. Data taken from analysis 5.

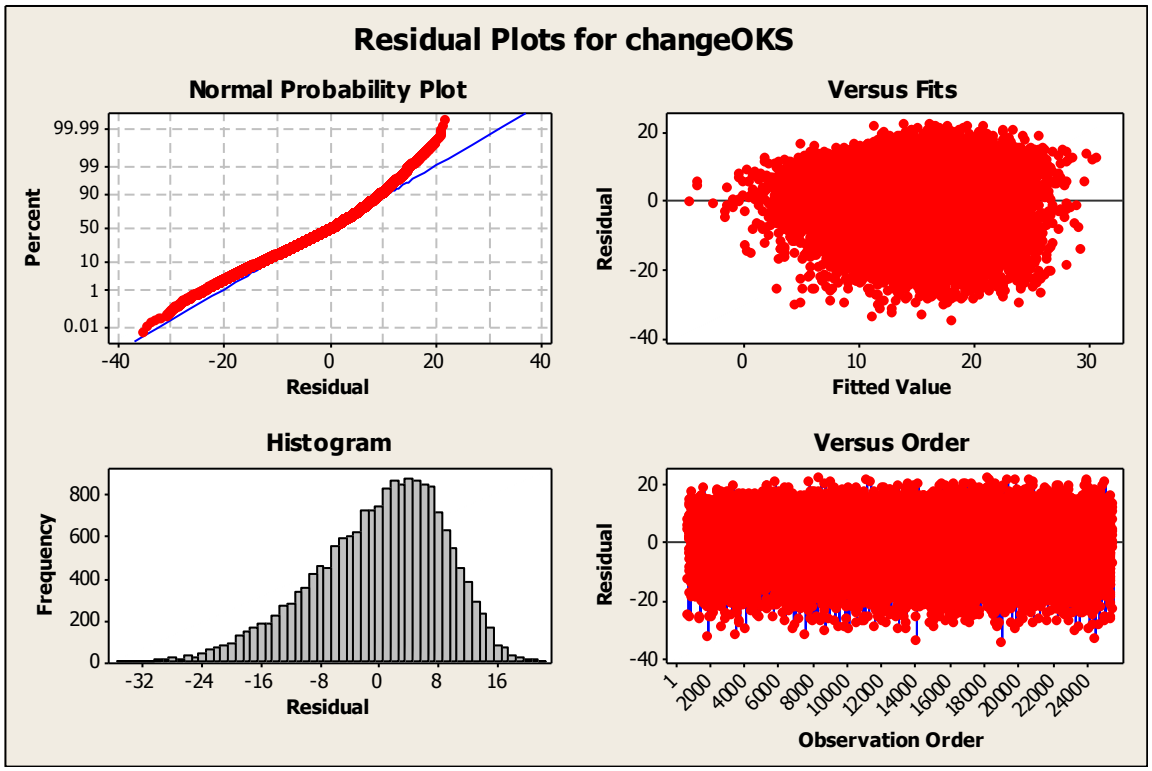


Figure 3.10: Example of standard 4 in 1 residuals plots used in the assessment of model adequacy. Data taken from analysis 1.

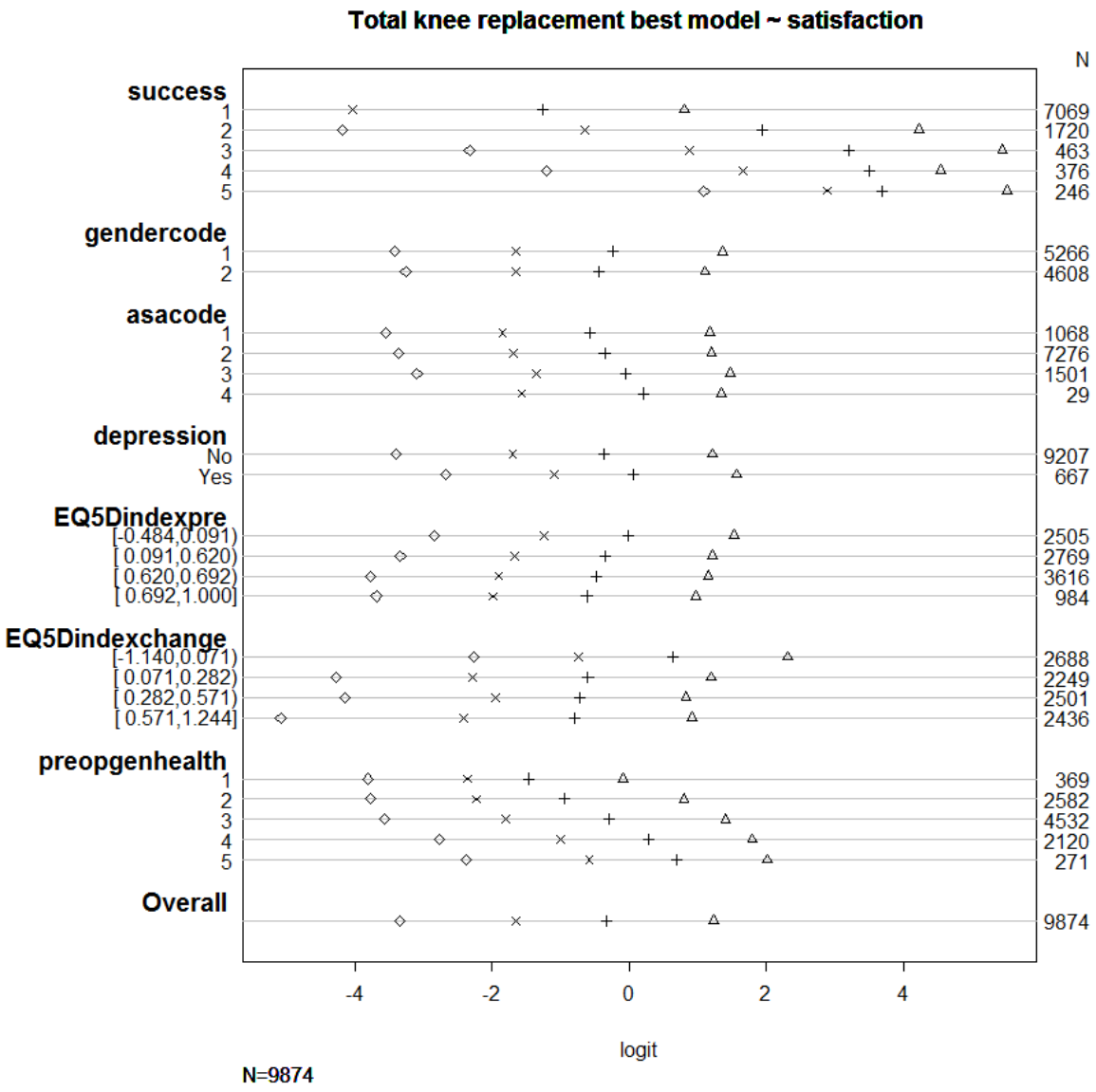


Figure 3.11: Example of the graphical plots used to assess the proportionality assumption as part of the ordinal regression modeling process. Data taken from analysis 6.

3.6.4 Structured equational modelling (SEM)

SEM investigates relationships among different processes by partitioning relationships among variables on the basis of a hypothetical pathway of interactions identified prior to analysis. It effectively challenges a prior hypothetical model of a system using observed data for that system. The paths between variables are defined in equation form with response variables related to two or more explanatory variables. The response variables in one equation then form the explanatory variables in other equations. SEM tests whether the variables in the path are interrelated by analysing their variances and co-variances.

Our SEM modeling procedure was first to fit a full model with all hypothesised pathways and then to remove all non-significant paths to create the simplest model containing only significant pathways. This was determined by two measures of model fit (Root Means Square Error of Association (RMSEA) and Goodness of Fit Index (GFI)) (Kline 2003) and the paths standard errors with boot-strapped 95% confidence limits. SEM is usually based on analysis of a covariance matrix, expressing how each of the drivers and outcomes in the system of interest co-vary. Since our data were a mixture of variable type (categorical, ordinal, continuous, binary) we used polychoric correlations to create a correlation matrix for analysis in the SEM. Polychoric correlations allow estimation of correlation between ordinal and (when extended to tetrachoric correlations) categorical variables. These correlations are rarely bivariate normal so this means that standard errors on parameter estimates derived from a SEM analysis are likely to be biased (Fox 2006). We therefore used a Bootstrapping approach to estimate means and standard errors for the SEM parameter estimates. Bootstrapping was undertaken by running the model 200 times with 200 separate sets of data. The rationale was that analysing the mean and standard deviations of the 200 random subsets of the data this would provide an unbiased estimate of the SEM parameter estimates. This non-normality meant that the standard errors of the parameter estimates for the SEM are unlikely to be reliable; accordingly we used Monte Carlo approaches to assess the SEM. Our rationale was if a model was adequate with a subset of the data then we would have very conservative estimates of the impacts of the individual pathways in the overall model.

3.6.5 Sample sizes

As part of their PROMs feasibility analysis Browne (2007) calculated the sample sizes needed for PROMs comparisons using estimates for the minimally important differences and standard deviations for the EQ5D and OKS. These are detailed in table 3.13 and are presented for a number of different power and significance levels. The sample sizes used in the current analyses were in excess of these minimum numbers suggested by Browne (2007).

	Distribution based			
	0.05 significance level		0.002 significance level	
	80% power	95% power	80% power	95% power
Oxford knee score	103	171	203	294
EQ5D (for knees)	53	87	104	150

Table 3.13: Distribution based sample size details for the Oxford Knee Score and the EQ5D for knees.

3.6.6 Statistical packages used for the current analyses

Statistical analysis was performed using Genstat 10th Edition (Lawes Agricultural Trust, Hemel Hempstead, UK), Minitab version 16 (Minitab Inc., Coventry, UK.) and the Statistical Package for the Social Sciences SPSS v19.0 (SPSS Inc., IBM Corporation, Armonk, New York). Structured equational modelling analysis was performed using the R statistical package (R Foundation for Statistical Computing, Vienna, Austria).

Chapter 4: Clinical outcomes after unicondylar knee replacement

4.1 The on-going debate about the role of the unicondylar knee replacement (UKR)

The number of unicondylar knee replacements (UKRs) performed in England and Wales continues to rise. In 2010, 6995 UKRs were registered with the National Joint Registry (NJR), a substantial increase from 2226 recorded in 2003, even allowing for the incomplete registration of patients at the inception of the NJR (NJR-AR 2012). Reports from specialist centres suggest that UKRs have clinical outcomes and survival equivalent to those of total knee replacement (TKR) when followed for 10 to 15 years (Svard 2001, Price 2005, Newman 2009, Pandit 2011). Independent analysis of the use of UKR in the United Kingdom NHS has suggested it to be cheaper, functionally superior and under-used (Willis-Owen 2009). However, while some non-specialist centres have been able to replicate the success of these specialist centres (Keys 2004, Rajesekhar 2004, Lisowski 2011), others have not (Fehring 2010, Mercier 2010, Dervin 2011).

The results reported from specialist centres are also not reflected in registry outcomes, where revision rates for UKRs are significantly higher than for TKRs (NJR-AR 2012, Koskinen 2008, Davidson 2009, Lidgren 2009). In the most recent NJR report the seven-year revision rate for UKR (16.6% (95% CI 15.3 to 18.1)) was markedly higher than that observed for cemented TKR (3.8% (95% 3.7 to 4.0)) (NJR-AR 2012). This finding mirrors the rates reported in the Australian registry, where the ten-year revision rates for UKR and TKR are 15.1% and 5.7%, respectively (Davidson 2009). Recent analysis from the Finnish registry reported that UKR was not cost-effective when the overall survival and cost of revision were considered at a national level (Koskinen 2008). National registry data also suggests that a unicondylar replacement converted to a total knee replacement performs poorly in comparison to a primary total knee replacement (Pearse 2010) and is no better in terms of survival than revision total knee replacement (Dudley 2008). This has led to additional concern over the higher revision rates observed for unicondlyar replacements. The performance of this implant in the hands of the ‘average’ orthopaedic surgeon and its wider use in the setting of the National Health Service therefore remains a matter of considerable debate.

Advocates of UKR defend the differences between UKR and TKR observed in registries by arguing that registries simply look at revision rate in isolation without considering the greater functional gains, which this implant type should theoretically produce (Goodfellow 2010, Cobb 2010). Following any operation, the patient's perception of the outcome is fundamental. Greater improvements in patient-reported outcome measures (PROMs), following UKR when compared to TKR may allay some of the concerns about implant survival and the cost-effectiveness of the former.

It has also been suggested that discrepancies in the revision rate may also be partly explained by differential revision thresholds for these types of implant (Goodfellow 2010, Cobb 2010), but nevertheless the data represent a reflection of the reality of current practice. The surgical threshold to proceed to revision may be influenced by the existing implant and potential surgical challenges faced. Such surgical selection bias is not currently accounted for when using revision as an outcome measure. Indeed, the use of revision as the sole endpoint has been challenged as unfairly discriminating against UKR for which excellent centre outcomes exist (Murray 1998, Keys 2004, Newman 2009, Lisowski 2011, Pandit 2011). Both Cobb (2010) and Goodfellow *et al* (2010) have highlighted the paradox observed in the data published from the New Zealand Registry (NZJR-AR 2010) which demonstrates equivalent excellent early and mid-term functional outcome scores for both implant types, but a marked difference in survivorship with consistently higher revision rates for UKR. Despite equivalent knee scores, a UKR was between four to six times more likely to be revised than a TKR with the same outcome. This suggests that UKR may be more susceptible to revision, especially in the case of implants presenting with unexplained pain (Cobb 2010, Goodfellow 2010).

A further criticism of registry analysis of UKR data has been a failure to examine medial and lateral UKR separately. Survival rates for UKR stratified by implant laterality are not currently available in any registry report, as implants are pooled together for the purpose of analysis.

The vast majority of UKR (>90%) are implanted in the medial compartment of the knee and if laterally inserted implants were to have a high failure rate the pooled analysis used by national registries may bias the results of medially inserted implants. This argument is supported by a number of small, historical studies which have reported high failure rates of 82% to 83% at five to ten years in patients undergoing lateral UKR (Heyse 2010, Gunther 1996, Ashraf 2002). However, recent publications have reported ten-year survival rates of 92% (Argenson 2008) and 98% (Lustig 2011), with some series reporting 100% survival at a mean of 5.2 years (Sah 2007) and 12.4 years, respectively (Pennington 2006). Therefore the true survival and the influence of lateral UKRs with registry analyses remains unclear.

UKR is a technically demanding procedure with a long learning curve (Hamilton 2010). This has led to concerns about how these implants perform in the hands of surgeons and centres that are unable to maintain their surgical experience as they perform small numbers of these procedures. The literature suggests that rates of revision are heavily dependent upon the reporting institution (Labek 2011), with the best results coming from institutions involved in the design of these implants and independent advocates of UKR (Murray 1998, Keys 2004, Newman 2009, Lisowski 2011, Pandit 2011), and the worst results coming from registry analyses (Koskinen 2008, Davidson 2009, Lidgren 2009, NJR 2012,). Revision rates from the best centres are comparable to TKR suggesting that higher operative volumes and surgical enthusiasm improve revision outcomes. This contention is supported by registry data from Sweden where lower rates of UKR revision were found in centres performing the highest operative volumes (>23/year) when compared to all other centres (Robertsson 2001). A similar association between revision rate and unit volume have also been reported for TKR (Katz 2004, Marlow 2010) and may be an indication of the standards of patient selection and operative performance employed in these centres (Robertsson 2001).

Ultimately these concerns need to be addressed to allow the debate regarding the use of UKR to move forward. The HTA funded Total or Partial Knee Arthroplasty Trial (TOPKAT) (HTA ref: 08/14/08, grant £2,700,878) is currently the only prospective nationally funded trial assessing the role of UKR. The trial aims to "*assess the clinical and cost effectiveness of*

Total Knee Replacements versus Unicompartmental Knee Replacements in patients with medial osteoarthritis". Recruitment has only recently commenced and it is not expected to publish until early 2020. It will therefore be a number of years before this trial can offer anything useful to this debate, a debate which may have moved on as newer implants and methods of surgery are developed. The scope of this trial is also limited as, due to its design and primary aim, it can only assess short and mid-term clinical outcomes (function and complications) and cannot assess longer term concerns such as the rate of revision and mode of failure. For the same reasons it will also be unable to assess the impact of variables such as surgeon volume upon implant survival and will be restricted to the analysis of medial UKR.

It may be possible to address a number of the issues raised above using registry data. The NJR and PROMs data available was therefore used in an attempt to answer a number of questions about UKR. These included:

1. Do patients undergoing UKR experience greater improvements in knee specific function and general health measures and lower rates of complications than patient undergoing TKR?
2. How do the reasons for revision differ between UKR and TKR and are more UKR revised for unexplained pain when compared to TKR
3. What proportions of all UKRs are implanted in the lateral compartment of the knee and are UKR survival influenced by implant laterality?
4. What patient and implant factors influence failure of medial and lateral UKRs and does their mode of failure differ?
5. How many different centres and surgeons are performing UKR and in what amounts?
6. Is there an association between centre and/or surgeon volume and revision rate for UKR?
7. If surgeons perform high volumes of UKR do they achieve revision rates comparable to those seen with cemented TKR?

These questions are addressed in the subsequent 4 analyses.

1. Analysis of functional improvements and rates of complications for UKR and TKR.
2. Analysis of the reasons for revision and rates of failure for unexplained pain for UKR and TKR.
3. Analysis of the reasons for revision and rates of failure for medial and lateral UKR.
4. Analysis of the effect of centre and surgeon operative volume on UKR rates of failure.

4.2 Analysis 1: Analysis of functional improvements and rates of complications for UKR and TKR

Analysis 1: Aim:

To determine if patients undergoing UKR experience greater improvements in knee specific function and general health measures and lower rates of complications than patient undergoing TKR.

Analysis 1: Methods:

This analysis was performed on the NJR-PROMs dataset described in section 3.4.4. All 23,393 TKR and 505 UKR were used for the analysis (table 3.7 section 3.4.4). PROMs questionnaires were collected at medians of 15 days pre-operatively (15 days (0 to 90) and 12 days (0 to 90) for TKR and UKR, respectively) and 199 days post-operatively (199 days (180 to 365) and 199 days (180 to 349) for TKR and UKR, respectively). Comparison of the demographics of the two groups demonstrates that there were marked differences between them, with the UKR patients being younger and in better overall health. The UKR group also had a greater proportion of males, operations performed in independent centres, and operations performed by a consultant (table 3.7 section 3.4.4).

Initial comparisons of the patient demographics and PROMs for the TKR and UKR groups was undertaken using a combination of independent-samples *t*-tests for continuous data, Fisher's and chi-squared tests for categorical data, and ordinal logistic regression for ordinal data. As with the PROMs feasibility pilot (Browne 2007) we aimed to compare the two implant types according to change in disease-specific PROM (OKS), change in generic PROM (EQ5D index), presented as both the EQ5D index value and in the tabular format recommended by the EuroQol group, and proportion of patients reporting any of the six recorded complications (see sections 3.5.3 and 3.5.5).

In order to allow a meaningful comparison between UKR and TKR it was important to account for differences in case mix (patients' characteristics), which could be potential confounding factors in any comparative analysis. We therefore sought to determine the case-mix variables which explained the largest proportion of the variance in each of the PROMs outcomes of interest. This was achieved using the standard modeling processes described in section 3.6.3. The models included all patient characteristics and clinical factors collected routinely by the NJR and pre-operative PROMs questionnaire pertinent to the comparison of UKR and TKR (table 3.8). In order to help limit the effects of multiple testing and ensure we had a model containing only the most important variables, the final models included only those variables that were significant at a p-value of < 0.001. Whereas body mass index (BMI) was significant ($p < 0.001$) in the models relating to the change in OKS and EQ5D index, there were sufficient concerns about the quantity of missing data (approximately 40%) and the possibility for recording bias that we excluded it from the model. Also, within the models BMI was only responsible for a small amount of the explained variance, and was correlated with other more influential variables responsible for a greater proportion of the explained variance in the model (pre-operative OKS: $r = -0.22$, $p < 0.001$; Age: $r = -0.25$, $p < 0.001$).

Using these models we were able to 'adjust' the observed PROMs outcomes taking into account the inherent background variability associated with each of the two types of implant. Adjustment was performed by comparing the differences in final model residuals for TKR and UKR. This effectively compares the differences in the residual variation associated with the two groups after the variation associated with the factors included in the final model has been accounted for. PROMs data for both the 'unadjusted' (direct reporting of the PROMs without accounting for case-mix variation) and 'adjusted' (PROMs adjusted for case-mix variation) outcomes for TKR and UKR were reported, so that the effects of adjustment for each outcome could be observed.

For the smaller UKR group, power calculations were undertaken to assess the size of effect that could be confidently detected, based on the sample size and standard deviations of the

change in OKS and EQ5D index. With a power of 80% and significance of p = 0.01 we could detect a difference in the change in OKS of 2.1 points and a difference in the change in EQ5D index of 0.07 points. These differences were below the effect sizes felt to be clinically significant for these two scores. Owing to the effects of multiple statistical testing within this analysis, borderline p-values between 0.01 and 0.05 should be interpreted with caution, and a p-value of < 0.01 was therefore used to indicate statistical significance. Model residuals and other checks of model adequacy were satisfactory and there were no significant 2-way interaction between variables included within the final models.

Analysis 1: Results:

Unadjusted PROMs: There were significant post-operative improvements in knee-specific and generic PROMs for both types of implant (table 4.1 and 4.2). Analysis of the unadjusted PROMs showed that the improvements in OKS and EQ5D were greater following TKR than following UKR, although the post-operative PROMs were better after UKR (table 4.1). These differences were largely explained by the differences in the pre-operative baseline scores for each of the two groups, emphasising the need to correct for these differences when analysing these data. The unadjusted overall rates of complications were similar for the two groups, although there was a trend for lower rates of wound complications (p = 0.02), lower rates of readmission (p = 0.02) and a higher rate of patients requiring further surgery (p = 0.02) in the UKR group (Table 4.3).

Outcome measure	TKR	UKR	p-value
Mean OKS (95% CI)			
Pre-operative	18.9 (18.8 to 19.0)	21.5 (20.8 to 22.2)	< 0.001
Post-operative	34.0 (33.9 to 34.2)	35.5 (34.5 to 36.4)	0.002
Change	15.1 (15.0 to 15.3)	13.9 (13.1 to 14.8)	0.007
Mean EQ5D (95% CI)			
Pre-operative	0.407 (0.403 to 0.411)	0.470 (0.442 to 0.497)	< 0.001
Post-operative	0.710 (0.707 to 0.714)	0.736 (0.711 to 0.760)	0.04
Change	0.303 (0.298 to 0.307)	0.266 (0.236 to 0.296)	0.02

Table 4.1: Comparison of the unadjusted patient reported outcome measures for total knee replacement (TKR) and unicondylar knee replacement (UKR) Key: OKS, Oxford knee score; EQ5D, EuroQol-5D.

	Pre-operative			Post-operative		
EQ5D dimension	TKR	UKR	p-value	TKR	UKR	p-value
Mobility (n, %)						
Level 1	1383 (*6*)	43 (*9*)	0.07	10 713 (*46*)	260 (*52*)	0.009
Level 2	21 582 (*92*)	455 (*90*)		11 909 (*51*)	224 (*44*)	
Level 3	47 (*0*)	0 (*0*)		31 (*0*)	0 (*0*)	
Not recorded	381 (*2*)	7 (*1*)		740 (*3*)	21 (*4*)	
Self-care (n, %)						
Level 1	15 934 (*68*)	369 (*73*)	0.06	17 839 (*76*)	400 (*79*)	0.04
Level 2	6917 (*30*)	122 (*24*)		4734 (*20*)	81 (*16*)	
Level 3	142 (*1*)	3 (*1*)		135 (*1*)	2 (*0*)	
Not recorded	400 (*2*)	11 (*2*)		685 (*3*)	22 (*4*)	
Usual activities (n, %)						
Level 1	2000 (*9*)	55 (*11*)	0.07	9554 (*41*)	238 (*47*)	0.03
Level 2	17 885 (*77*)	391 (*77*)		11 958 (*51*)	228 (*45*)	
Level 3	3065 (*13*)	50 (*10*)		1104 (*5*)	20 (*4*)	
Not recorded	443 (*2*)	9 (*2*)		777 (*3*)	19 (*4*)	
Pain/discomfort (n, %)						
Level 1	224 (*1*)	6 (*1*)	< 0.001	7294 (*31*)	188 (*37*)	0.01
Level 2	13 604 (*58*)	343 (*68*)		13 850 (*59*)	263 (*52*)	
Level 3	9003 (*39*)	144 (*29*)		1337 (*6*)	31 (*6*)	
Not recorded	562 (*2*)	12 (*2*)		912 (*4*)	23 (*5*)	
Anxiety/depression (n, %)						
Level 1	14 343 (*61*)	312 (*62*)	0.67	17 270 (*74*)	362 (*72*)	0.30
Level 2	7652 (*33*)	162 (*32*)		4783 (*20*)	111 (*22*)	
Level 3	798 (*3*)	21 (*4*)		509 (*2*)	8 (*2*)	
Not recorded	600 (*3*)	10 (*2*)		831 (*4*)	24 (*5*)	

Table 4.2: Comparison (%) of the pre- and post-operative responses for each of the five EuroQol (EQ5D) domains (Level 1, no problems; Level 2, moderate problems; Level 3, severe problems; NR, not recorded; TKR, total knee replacement; UKR, unicondylar knee replacement).

Post-operative complications (n, %)	TKR (n = 23,393)	UKR (n = 505)	p-value
Allergy	3870 (*17*)	65 (*13*)	0.03
Urinary tract infection	2525 (*11*)	50 (*10*)	0.52
Bleeding problems	1533 (*7*)	27 (*5*)	0.28
Wound problems	2808 (*12*)	43 (*9*)	0.02
Readmission	2290 (*10*)	33 (*7*)	0.02
Further surgery	848 (*4*)	28 (*6*)	0.02
Patients stating ≥ 1 of the above complications (n, %)	8985 (*38*)	177 (*35*)	0.12

Table 4.3: Comparison of patients' reported post-operative complications following total knee replacement (TKR) and unicondylar replacement (UKR).

Development of risk-adjustment model: Regression models were generated for each of the response variables of interest (change in OKS, change in EQ5D index, and rate of complications). Initial univariate analysis of the effect of the implant type on the PROMs outcomes showed there was a relationship for the change in OKS ($p = 0.007$) and EQ5D index ($p = 0.02$) but not the rate of complications ($p = 0.12$). This is consistent with the preliminary unadjusted analysis. However, once all other variables were entered into the model the effect of implant type diminished and was no longer significant.

The variables included in the final models are given in table 4.4. For the change in OKS and EQ5D index, the most important variable in the models was the relevant pre-operative score. In both cases there was a greater change in score as the pre-operative score decreased, so that patients who were worst to begin with had the best improvement. This highlights the influence of the baseline score on the ability to improve post-operatively, and the possible ceiling effects of these two scores, whereby patients with better pre-operative scores are unable to improve to the same extent as those with poorer scores, owing to the inability of these scores to detect top-end differences. This further emphasises the importance of appropriate adjustment when comparing PROMs. Two variables appeared in all three models: pre-operative general health and anxiety level. For these variables, along with pre-operative disability, number of comorbidities and depression, the observed relationship with the response variable was for smaller improvements in score with greater levels of anxiety, depression, pre-operative disability and worsening pre-operative general health.

Adjusted PROMs: After adjustment, the mean difference between the change in OKS for the TKR and UKR groups was 0.0 (95%CI -0.9 to 0.9; $p = 0.96$). The adjusted change in EQ5D index was similar for both UKR and TKR (mean difference of 0.09 (95%CI -0.015 to 0.034; $p = 0.37$) (table 4.5). The odds of developing a complication were also similar for the two groups (odds ratio (OR) TKR *versus* UKR = 1.17 (95% CI 0.92 to 1.37); $p = 0.24$).

	Response variable		
Variables included in the final risk-adjustment models	Change in OKS	Change in EQ5D	Complication rate
Age	X	X	
Gender		X	X
Anxiety level	X	X	X
Depression		X	
Hospital type	X		
Indication for surgery	X		
Number of comorbidities	X		
Pre-operative disability	X	X	
Pre-operative general health	X	X	X
Pre-operative OKS	X	X	
Pre-operative EQ5D index		X	
Pre-operative EQ5D health VAS		X	X
Variance explained by the risk adjustment models (%)			
Full model (all 19 variables)	*22*	*52*	-
Chosen model (variables listed above)	*21*	*51*	-
Pre-operative PROM of response variable	*15*	*45*	-

Table 4.4: The risk adjustment models for the change in Oxford knee score (OKS) (multiple regression model), change in EuroQol (EQ5D) index (multiple regression model) and complication rate (logistic regression model). VAS, visual analogue score. Variables included were significant at $p < 0.001$.

	TKR *versus* UKR	p-value
Change in OKS		
Unadjusted difference	+1.2 (95% CI 0.3 to 2.1)	0.007
Adjusted difference	0.0 (95% CI -0.9 to 0.9)	0.96
Change in EQ5D		
Unadjusted difference	+0.037 (95% CI 0.006 to 0.067)	0.02
Adjusted difference	-0.009 (95% CI -0.034 to 0.015)	0.37
Complications		
Unadjusted odds ratio	1.16 (95% CI 0.96 to 1.39)	0.12
Adjusted odds ratio	1.17 (95% CI 0.92 to 1.37)	0.24

Table 4.5: Unadjusted and adjusted mean differences in patient reported outcome measures for the total knee replacement (TKR) and unicondylar knee replacement (UKR) groups. For change in Oxford knee score (OKS) and EuroQol (EQ5D), positive differences favour TKR, and negative differences favour UKR. For complications odds ratios are given for TKR relative to UKR.

Analysis 1: Discussion

This analysis has shown that although unadjusted improvements in PROMs were greater for TKR than for UKR, once adjustments had been made for differences in case mix and baseline scores, there were no statistically significant or clinically important differences in the improvements in either knee-specific, generic outcomes or complication rates between these two implant types.

There is only one other registry study comparing clinical outcomes of UKR and TKR. Lygre *et al* (Lygre 2010) interrogated the Norwegian registry and compared 372 UKRs and 972 TKRs at a mean of 6.5 years post-operatively. Their analysis included only post-operative knee injury and osteoarthritis outcome (KOOS) (Roos 1998) and EQ5D scores, and comparisons with pre-operative scores were not possible. Whereas some statistically significant differences were found between UKR and TKR for the KOOS subscales of symptoms, function in sport and recreation and function in daily living, these differences decreased with adjustment, and were below the accepted level for a minimal perceptible clinical difference in outcome. They therefore concluded that the two implants demonstrated similar levels of pain and function and questioned the use of UKR, given the reported higher rates of revision (Lygre 2010). Both the NJR and the Swedish registry investigated patient satisfaction following knee arthroplasty (Robertsson 2000, Baker 2007). The Swedish study found no difference between the implant types for general satisfaction (Robertsson 2000), whereas the NJR study found a higher level of satisfaction after TKR, despite the UKR group having a better post-operative OKS (Baker 2007).

As in this analysis, Pearse *et al* (2010) found in their analysis of the New Zealand registry that the unadjusted mean post-operative OKS was better for UKR (39.2) than for TKR (37.2), but they provided no further analysis nor attempted to account for the potential confounding differences in case-mix. The clinical results of Newman, Pydisetty and Ackroyd (Newman 2009) also suggest that there are no significant differences in pain scores between TKR and UKR in the longer term. Pandit *et al* (2011) reported on the first 1000 phase-3 Oxford UKRs

from the originating centre. The mean OKS at five years was 41.3 (SD 7.2), compared with 24.7 (SD 8.7) pre-operatively, a mean rise of 16.6 points. These excellent results are superior to those in this analysis, where we found a mean unadjusted OKS improvement of 13.9 for UKR, suggesting that the Oxford experience cannot be extrapolated to the national scale.

The final outcome of interest was the overall complication rate, which was very high for both groups. This was probably due to the fact these were patients' reports rather than verified hospital data, and included complications such as allergy and urinary infection, which cannot be directly attributed to the implant type and may be difficult for patients to define. Whereas there was no difference in the overall complication rates, there was variation between the implants in the rates of wound infection, readmission and reoperation.

Analysis 1: Summary

This analysis has shown that there are no differences in the improvements in either knee-specific or general health PROMs between UKR and TKR in a large cohort of registry patients. Given the on-going concerns about the significantly higher revision rates for UKR observed in registries worldwide, this information about the expected improvements in function for UKR relative to TKR should be of prime interest to the orthopaedic community.

Analysis 1: Limitations and Reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

While it was possible to adjust the UKR and TKR groups for differences in a range of patient characteristics one key variable for which we were unable to correct for was the severity and distribution of pre-operative arthritis. UKR are designed for people for isolated medial or

lateral compartment disease and are not used in patients in whom both of these compartments are involved. In contrast, TKR can be used for patients with either medial or lateral or combined disease. The cohorts of patients being compared were therefore not directly comparable. This is important as the 'real world' clinical problem relates to which treatment is best for isolated medial or lateral compartment disease. There is wide agreement that the best treatment for combined compartment disease is TKR and so these patients should ideally be excluded from any comparison of the two procedures. This was not possible due to the confines of the current datasets. The models developed for analysis 1 included variables such as age, pre-operative OKS and general health, which were correlated with disease severity and were probably acted as surrogate markers for it. This allowed us to, in part, adjust for differences in the severity of disease between the groups and gain the best possible representative comparison of these implants using registry data. However despite the inclusion of these variables the data we have presented is only able to give an indication of the answer and does not definitively answer the question posed. Capture and inclusion of additional data (e.g. radiographic data) may have overcome some of these problems as it would allow specific groups of patients (e.g. patients with isolated medial compartment disease) to be identified and directly compared. However, as with all observational research a lack of information about the clinician's rationale for choosing one implant over another would remain. The only way to adequately answer this question is by using defined inclusion / exclusion criteria and randomisation to take the surgical decision out of the hands of the clinician as is being done in the TOPKAT Trial.

This analysis revealed a possible issue relating to the representativeness and coverage of the PROMs data collection, specifically for UKR procedures. The numbers of UKR in this analysis were smaller than might be expected (505 of 23,898 (2%)), given that UKRs represent approximately 8% of all primary knee arthroplasties performed in England and Wales (NJR-AR 2012). The reasons for this are unclear but may in part relate to the observation that the collection of PROMs data is currently only from NHS-funded operations, and that more UKRs than TKRs are privately funded and performed in non-NHS hospitals. While the patient characteristics were similar, this observation raises the question of whether the surgeon,

centre and operative characteristics of these procedures are representative of national practice.

Despite the large numbers of explanatory variables available to us our models only explained a small proportion of the variation in the PROMs improvements for both the OKS (full model 22%) and EQ5D (full model 52%). This meant that between 48% and 78% of the variation in these outcomes was unexplained. A proportion of this unexplained variation would be due to the 'natural' variation observed with any outcome measure but it raises the concern that a variety of other unmeasured variables are influencing these outcomes and acting as sources of confounding. Variables such as mental health scores, education status and patient expectation are known to influence functional outcome after knee replacement and were not available for this analysis. It is possible that if these variables were available we could have produced more robust models or possibly that we might have observed a difference in outcome between the UKR and TKR procedures.

This comparison might also have been limited by the functional outcome measures chosen by the PROMs project. Advocates of UKR believe this procedure is better at restoring high end function, and cite this as a benefit of this procedure. Due to the ceiling effect observed with the OKS this score may not detect differences at the top end of the score, between the highest functioning individuals. This may be a source of bias against UKR procedures which could be reduced if different outcome measures were employed.

Analysis 1 was based on functional data from all the UKR for whom PROMs records were available. Later analysis (analysis 3) determined the laterality (medial *versus* lateral) of the NJR UKR cohort, allowing us to also determine the laterality of the NJR-PROMs UKR cohort. This revealed that there were 484 medial and 21 lateral UKR within the NJR-PROMs UKR group. As the number of lateral UKRs were small we were unable to determine whether the lack of difference in the PROMs between the medial and lateral UKRs was a valid observation, or simply a function of low statistical power (type II error). Given this

uncertainty it would probably have been better if the lateral UKR had been excluded and the analysis had been based exclusively upon the medial UKRs.

If it were to be repeated, this analysis would benefit from the inclusion of radiological data to ensure groups were matched for disease severity, inclusion of a greater number of explanatory variables to assist with the creation for more robust models, restriction of the UKR group to include only medial UKR and the use of outcomes measures that are better suited to detecting differences in high end function. Trying to determine the relative merits of UKR and TKR based on observational data is always likely to encounter methodological limitations and the best way to address this problem is through pragmatic well constructed clinical trials such as the TOPKAT trial.

4.3 Analysis 2: Analysis of the reasons for revision and rates of failure for unexplained pain for UKR and TKR

Analysis 2: Aim:

To determine how the reasons for revision differ between UKR and TKR and establish whether more UKR revised for unexplained pain when compared to TKR.

Analysis 2: Methods:

The NJR data set described in section 3.4.2 was used for this analysis. All 366,965 TKR and 35,749 UKR were included in the analysis irrespective of the indication for surgery. Demographic details for the TKR and UKR cohorts are given in table 4.6. Significant differences were seen for a number of the demographic variables reflecting the differing populations that undergo TKR and UKR respectively.

As of December 2010, 6,075 implants had undergone a revision procedure according to registry records (figure 3.1 section 3.4.2). For those cases that underwent revision both the institution performing the revision and the reason for revision were available for analysis. To overcome the problem of multiple reasons for revision we employed the hierarchical strategy for determining the primary reason for revision described in section 3.5.2 (table 3.9). This meant the primary reason for revision could be determined for all TKR and UKR revisions.

Group	TKR	UKR	p value
PRIMARY SURGERY			
Number	**366,965**	**35,749**	
Mean Age (Years)	70.4 (S.D 9.2)	64.5 (S.D 9.6)	<0.001
Gender (Female:Male)	210,769F:156,196M	17,231F:18,518M	<0.001
Primary indication osteoarthritis	355,692 (97%)	35,345 (99%)	<0.001
REVISION SURGERY			
Number	**4503**	**1572**	
Mean Age (Years)	68.2 (S.D 9.4)	64.0 (S.D 9.8)	<0.001
Gender (Female:Male)	2371F:2132M	806F:766M	0.34

Table 4.6: Summary of the primary and revision demographics for the unicondylar and total knee replacement groups.

After determining the primary reason for each revision we specifically analysed the proportion of TKR and UKR undertaken for unexplained pain to see if there was any support for the supposition of Cobb (Cobb 2010) and Goodfellow (Goodfellow 2010) that there is a differing threshold for revision in the face of the implant presenting with unexplained pain. The two implant types were compared using Cox regression to adjust for differences in age, gender, American Society of Anaesthesiologists (ASA) grade and indication for primary surgery. Analyses were based on three different endpoints: revision for all indications; revision for unexplained pain; and revision for any other reason as endpoints. A p value of <0.05 was deemed to be statistically significant. For the Cox regression analysis proportionality assumptions were broadly met.

Analysis 2: Results

A breakdown of the primary reason for revision stated on the data collection form for both groups are given in table 4.7. The percentage of patients revised for unexplained pain was 23% (364 of 1572 revisions) in the UKR compared to 9% (408 of 4503) in the TKR group (p<0.001). The corresponding odds of revision for unexplained pain was 3.0 (95%CI 2.6 to 3.5) greater for UKR when compared to TKR. Of interest is the finding that, of the UKR revised, only 67 patients (4%) had arthritis progression as an indication for revision which is lower than previously reported (Gioe 2003). This may, however, be a reflection of the modes

of failure observed with shorter term follow up where progression of arthritis does not have a chance to develop, combined with the observation that progression of arthritis was only introduced as a specific option on the revision data collection form in later versions (1056 of 1572 revisions).

Reason for Revision	**TKR n=366,965**		**UKR n=35,749**	
	Number of revisions (%)	% risk of revision	Number of revisions (%)	% risk of revision
Infection	1295 (29%)	0.35%	90 (6%)	0.25%
Malalignment	382 (8%)	0.10%	86 (5%)	0.24%
Loosening / Lysis	1069 (24%)	0.29%	477 (30%)	1.33%
Component Dissociation	35 (1%)	0.01%	44 (3%)	0.12%
Component wear / Polyethylene wear	73 (2%)	0.02%	29 (2%)	0.08%
Implant breakage	15 (0%)	0.00%	2 (0%)	0.01%
Dislocation / Instability	579 (13%)	0.16%	144 (9%)	0.40%
Fracture of bone	75 (2%)	0.02%	36 (2%)	0.10%
Progression of arthritis / disease	34 (1%)	0.01%	67 (4%)	0.19%
Arthrofibrosis / Stiffness	265 (6%)	0.07%	22 (1%)	0.06%
Unexplained pain	*408 (9%)*	0.11%	*364 (23%)*	1.02%
Other (if not listed above) / No reason	273 (6%)	0.07%	211 (13%)	0.59%
Total	**4503 (100%)**	1.23%	**1572 (100%)**	4.40%

Table 4.7: Frequency of each reason for revision for the total and unicondylar knee replacement groups. In cases where multiple reason were stated the dominant reason for revision was determined using the hierarchy in Table 3.9 (section 3.5.2).

The mean time to revision was similar for the two groups with an overall mean of 2.0 years (UKR 2.0 years, TKR 2.0 years, (p= 0.19)). The mean time to revision was not different when unexplained pain was stated as the reason for revision (UKR 2.1 years, TKR 2.0 years, (p= 0.61)). In total 295 centres and 570 surgeons performed revisions of failed UKR implanted within the lifespan of the registry (April 2003 to December 2010). The median number of UKR revised per centre was 3 (Range 1 to 56) and per surgeon was 2 (Range 1 to 35). The centres and surgeons performing the lowest numbers of revisions were those that performed the lowest numbers of primaries. In comparison to TKR, UKR revisions were more likely to be performed in the same institution that undertook the primary procedure (UKR 1162 of 1572 revisions (74%), TKR 3047 of 4503 revisions (68%) (p<0.001)) and be performed by the same surgeon (UKR 1019 of 1572 (65%), TKR 2189 of 4503 (49%), p<0.001).

Kaplan Meier survival curves for UKR and TKR using the endpoints of all revisions, revision for unexplained pain and revisions for all other reasons are given in figures 4.1, 4.2 and 4.3. At 5 years the failure rates for all revisions (UKR 6.9% and TKR 2.0%) and revisions for unexplained pain (UKR 1.6%, TKR 0.2%) were significantly greater following UKR (both p<0.001). After adjustment using Cox regression the risk of revision was greater for UKR for all revisions (Hazard Ratio (HR) = 2.82 (95%CI 2.66 to 2.99, p<0.001) and revisions for unexplained pain (HR = 6.76 (95%CI 5.84 to 7.83), p<0.001). After removal of those replacements revised for unexplained pain the risk of revision for all other reasons was still greater for UKR (HR = 2.39 (95%CI 2.27 to 2.56), p<0.001) suggesting proportionately more UKR were also revised for other reasons.

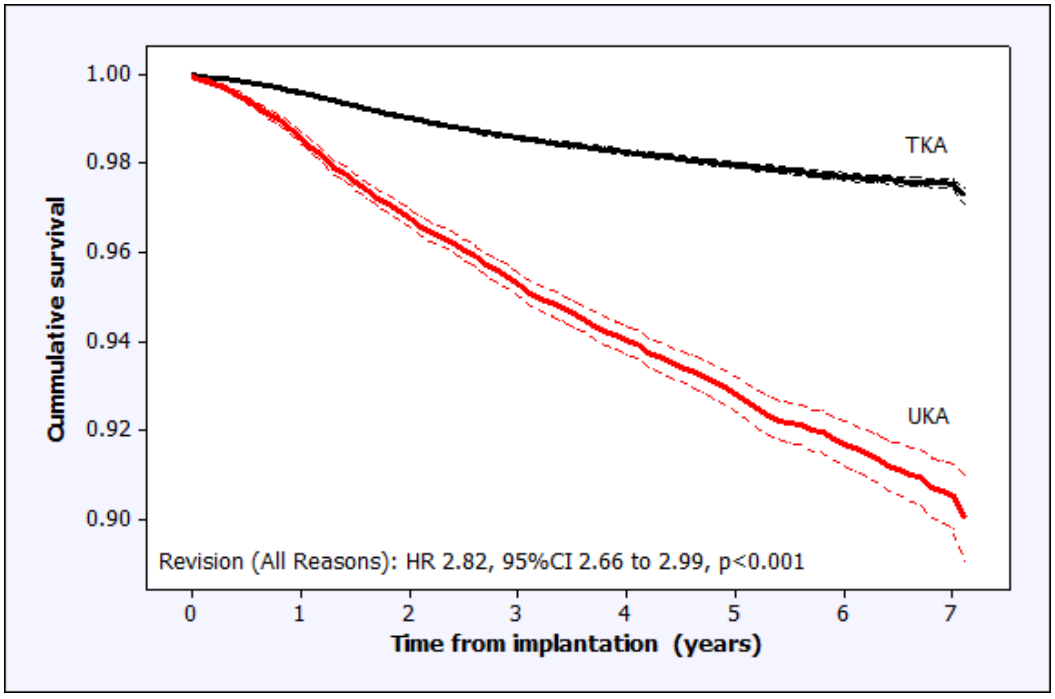


Figure 4.1: Kaplan-Meier survival plot for unicondylar (UKA) and total knee replacement (TKA) groups: Endpoint all revisions. Analysis adjusted for age, gender, ASA grade and operative indication.

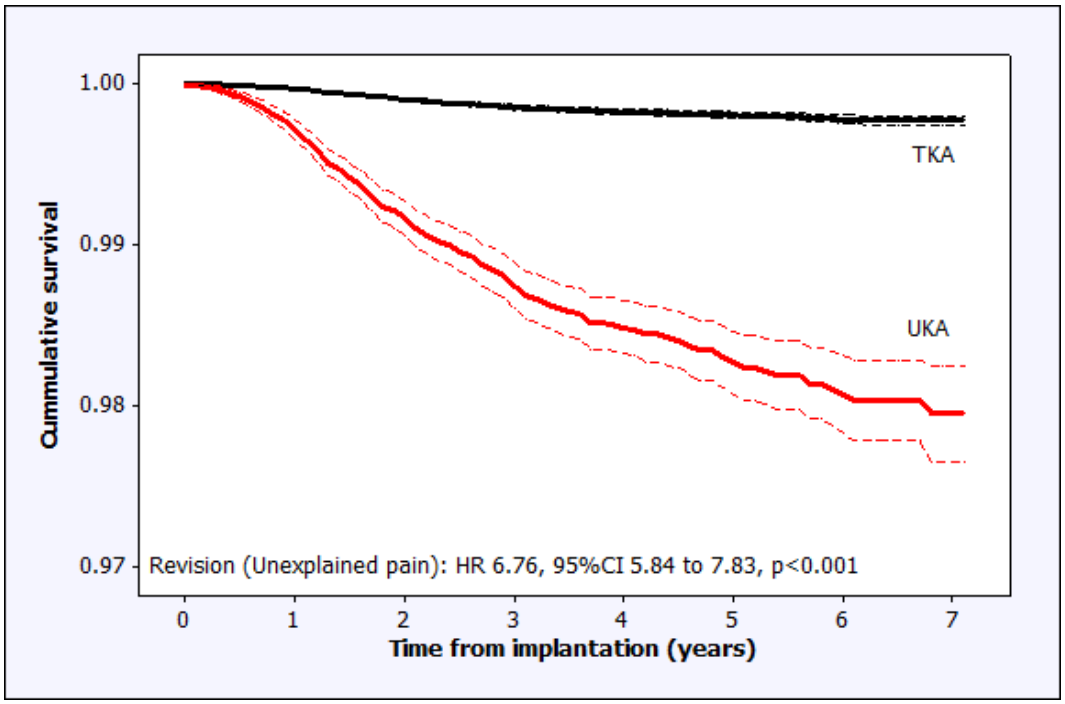


Figure 4.2: Kaplan-Meier survival plot for unicondylar (UKA) and total knee replacement (TKA) groups: Endpoint revision for unexplained pain. Analysis adjusted for age, gender, ASA grade and operative indication.

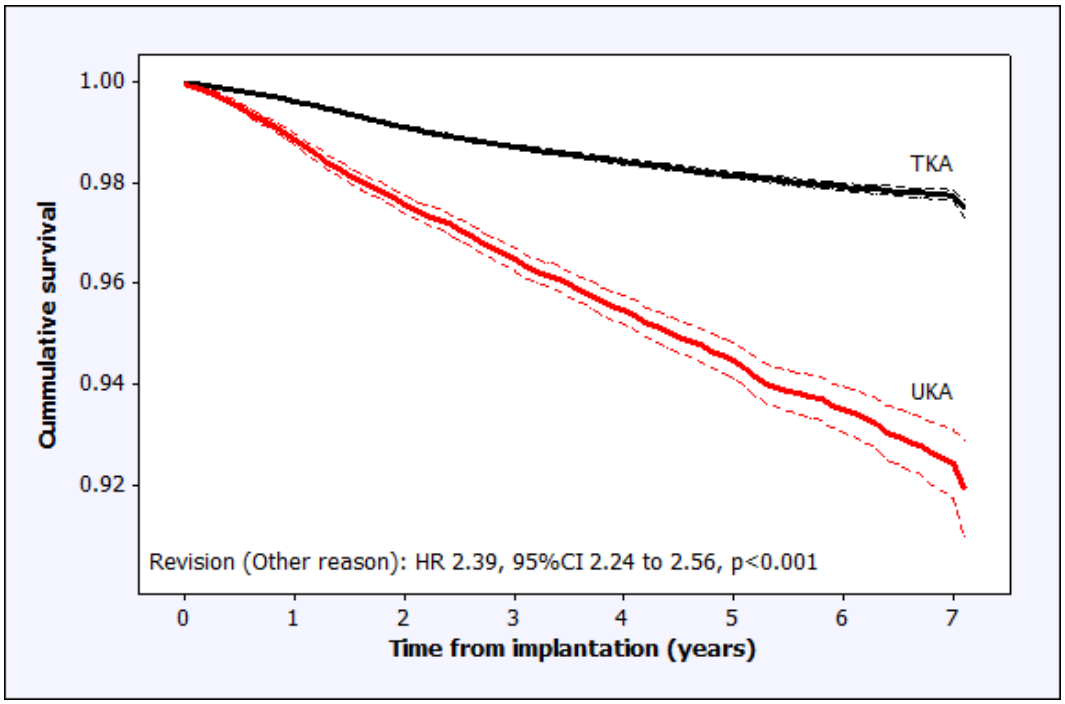


Figure 4.3: Kaplan-Meier survival plot for unicondylar (UKA) and total knee replacement (TKA) groups: Endpoint revision for all other reasons. Analysis adjusted for age, gender, ASA grade and operative indication.

Analysis 2: Discussion

This analysis found that the risk of revision for unexplained pain is greater following UKR and that proportionally more UKR were revised for this reason. These findings mirror those from other registries in which data on "pain" as a reason for revision is available (AJR-AR 2010, NZJR-AR 2010). The proportion of revisions for "pain" is greater after UKR in the registries of New Zealand (NZJR-AR 2010) (121 of 284 UKR revisions (43%) *versus* 271 of 835 TKR revisions (32%, p=0.002) and Australia (AJR-AR 2010) (331 of 2882 UKR revisions (12%) *versus* 736 of 8155 TKR revisions (9%) (p<0.001). The reported percentages vary between registries because of different methods of data collection. Previous analysis from the Norwegian Registry is limited by a small sample size but also suggests that UKR are more commonly revised for unexplained pain (Furnes 2007). The findings of this analysis are specific to the cause of revision and therefore clarify the consistent finding in all registries that UKR is more often revised for unexplained pain.

Reliance solely upon revision rate as an endpoint, whilst a definitive event, is highly subjective (Wylde 2011). Surgeon preference and the perceived ease of revision influence the decision to revise a particular implant type. It is therefore important to consider the reason for revision alongside the rate of revision if the behaviour of specific implant types is to be understood. Revision for unexplained pain represents a relatively small proportion of all revisions and once it is accounted for, there still remains a much greater revision rate for UKR from other causes when compared to TKR. It therefore seems unlikely that a low threshold for revising the implant with unexplained pain is biasing the overall revision rate for UKR.

It has been suggested that, in the hands of less experienced surgeons, there is a tendency towards earlier revision for UKR when compared to an equivalently performing TKR (Goodfellow 2010). This could be an additional source of bias against UKR. While the majority of UKR revisions were carried out by centres and surgeons which may be considered infrequent revisers the time to revision for UKR and TKR revised for unexplained

pain were equivalent. This finding is also supported by data from the New Zealand registry (NZJR-AR 2010).

There are a number of potential explanations for these findings. Firstly, if UKR revision is perceived as an easy procedure, with likely benefit to the patient, surgeons may have a lower threshold for revising the UKR with unexplained pain. Secondly, there may be more patients with a painful UKR than those with a painful TKR. Thirdly, UKR may produce less consistent outcomes than TKR, meaning they are either very good or very bad. This might explain the paradox between statistically equivalent functional outcomes, but higher revision rates with UKR. Fourthly, unexplained pain may be caused by subtle problems that the surgeon is unable to detect and/or document on a standardised form such as that used by the NJR.

Analysis 2: Summary

This analysis demonstrates that unexplained pain is a reason for revision more commonly with UKR than TKR. The reasons for this remain unclear. It suggests that using revision rates in isolation to compare these two implants is misleading. This supports the conclusions of others that more detailed analysis is required when using registry data to compare implants (Goodfellow 2010, Cobb 2010, Horan 2010).

Analysis 2: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

The reason for revision is related to the time from primary surgery. Early revisions performed in the first year following primary surgery are more likely to be due to infection

or issues relating to component malalignment. Late revisions performed more than 10 years following primary surgery are more likely to be due to loosening, lysis and component wear. As such the distribution of reasons for revision will change dependent upon the duration of follow up. Currently the NJRs follow up is limited to 7 years. As the registry matures it is likely that the distribution in the reasons for revision will change. Comparison of the reasons for revision for the UKR and TKR groups is valid as the duration of follow up is similar for both. However, one must be careful when interpreting the number of revisions and associated risk of revision for each individual group as these may change with increased follow up.

The hierarchy used within this analysis is based upon the hierarchy used by the Australian Joint Registry for their registry analyses (AJR-AR 2012). This hierarchy is based on expert consensus without supportive evidence. As such one could make a case for an alternative hierarchy which could alter the observed numbers of revisions within each group. This is particular important when one considers that, for the majority of cases (approx. 60%), two or more reasons for revision were recorded on the NJR K2 form. Alteration of the hierarchy therefore has the potential to significantly alter the results. This problem could be resolved if the K2 form were to explicitly ask the surgeon what they felt to be the main cause for revision, as this would allow one primary cause to be linked to each revision procedure.

This analysis was based on information recorded about the reason for revision. No information was available relating to the surgical decision making process. We therefore have no way of adequately determining why patients underwent revision and whether there was a differing threshold to revision based upon the primary implant type. The UKR group was younger than the TKR group. This difference could have had an equal if not greater bearing on the decision to proceed to revision. We have therefore only been able to surmise that the discrepancy in revision rate for UKR and TKR is not related to a differing threshold for revision. This is based on the observation that once revisions for unexplained pain are excluded the number of revisions performed for more objective reasons (e.g. loosening, infection, wear, instability etc.) were still significantly greater in the UKR group. This suggests

that more UKR fail in a manner that necessitates and justifies revision irrespective of the revising surgeon feelings about the ease of revision. This is, however, just one possible explanation for the results observed in this analysis and is by no means conclusive.

It is difficult to imagine how any trial could adequately assess something as subjective as the decision to proceed to revision surgery. As such it is unlikely that this argument will be resolved in the near future.

4.4 Analysis 3: Analysis of the reasons for revision and rates of failure for medial and lateral UKR

Analysis 3: Aims

To determine:

1. What proportion of all UKRs are implanted in the lateral compartment of the knee.
2. Whether UKR survival is influenced by implant laterality.
3. Which patient and implant factors influence failure of medial and lateral UKR.
4. Whether the mode of failure differs between medial and lateral UKR.

Analysis 3: Methods

The 35,749 UKR recorded in NJR dataset described in section 3.4.2 were used as the basis for analysis. For each record an attempt was made to determine whether the implant was used in the medial or the lateral compartment by combining information for the side of operation with details held within the description of the components. The descriptions were assessed for keywords, including 'RIGHT MEDIAL/LEFT LATERAL', 'RIGHT', 'RM/LL', 'R MDL/ L LAT', 'MRT/LLT' and 'RM', which specified that the component was designed for use in the medial side of a right knee or the lateral side of a left knee. Similar keywords were sought for components designed for use in the medial side of a left knee or the lateral side of a right knee. This information was then used alongside the side of operation (right or left knee) to establish whether the implant was used in the medial or the lateral compartment of the knee.

From the original cohort, 125 UKRs were excluded as they could not be accurately classified. In total, 20 different brands of UKR were implanted, and for 13 of these laterality could be determined from either the femoral or the tibial component description. Four of the remaining seven brands totaled only 21 UKRs, and the other three totaled 1649 (AMC

Uniglide (Corin, Cirencester UK)), 625 (Sled (Waldemar Link, Hamburg, Germany)) and 219 (UC-PLUS (Endo Plus, Rotkreuz, Switzerland)) implants. For these latter three brands the manufacturers were contacted to determine whether they could provide additional information that would allow component laterality to be determined. Both Corin and Waldemar Link confirmed that no such information was available. In addition, the laterality of 263 (1%) of the 26,294 Oxford UKRs (Biomet, Warsaw, Indiana, USA) could not be determined because of unique component descriptions. In total, laterality could therefore be determined for 32,847 of 35,624 UKRs (92.2%) (figure 4.4)

Using revision for any reason as the endpoint, implant survival was calculated for both the medial and lateral UKR groups using Kaplan-Meier and life table analysis. Survival rates were compared using the log rank (Mantel-Cox) test. Supplementary comparison of the medial and lateral groups was performed after adjustment for any differences in patient age group (< 55 *versus* 55 to 65 *versus* 65 to 75 *versus* > 75 years), gender, American Society of Anesthesiologists (ASA) grade, indication for surgery (osteoarthritis/other) and implant design (fixed bearing with all-polyethylene tibial component/fixed bearing with metal-backed one-piece tibial component/fixed bearing with modular tibial component/mobile bearing with modular tibial component) between the groups using Cox's proportional hazards method. This method was then employed to determine which of these covariates influenced the rates of failure for medial and lateral UKRs separately. These covariates were chosen as they mirror the factors used by other registries when stratifying the risk of revision for specific types of implant (Davidson 2009, Lidgren 2009). The design of the implant was included as a covariate as there was interest in establishing whether the performance of different bearing types and tibial designs varied according to laterality. Information relating to design was obtained from the component codes and descriptions for all but 947 medial and 95 lateral implants. Information for all of the other covariates was available for all implants in both the medial and the lateral groups.

Finally the reasons for revision for both medial and lateral UKR were examined using the hierarchy described in section 3.5.2 (table 3.2). The analysis also considered the effect of

bearing type (fixed *versus* mobile) on the reason for revision within each group. A p-value <0.05 was considered statistically significant.

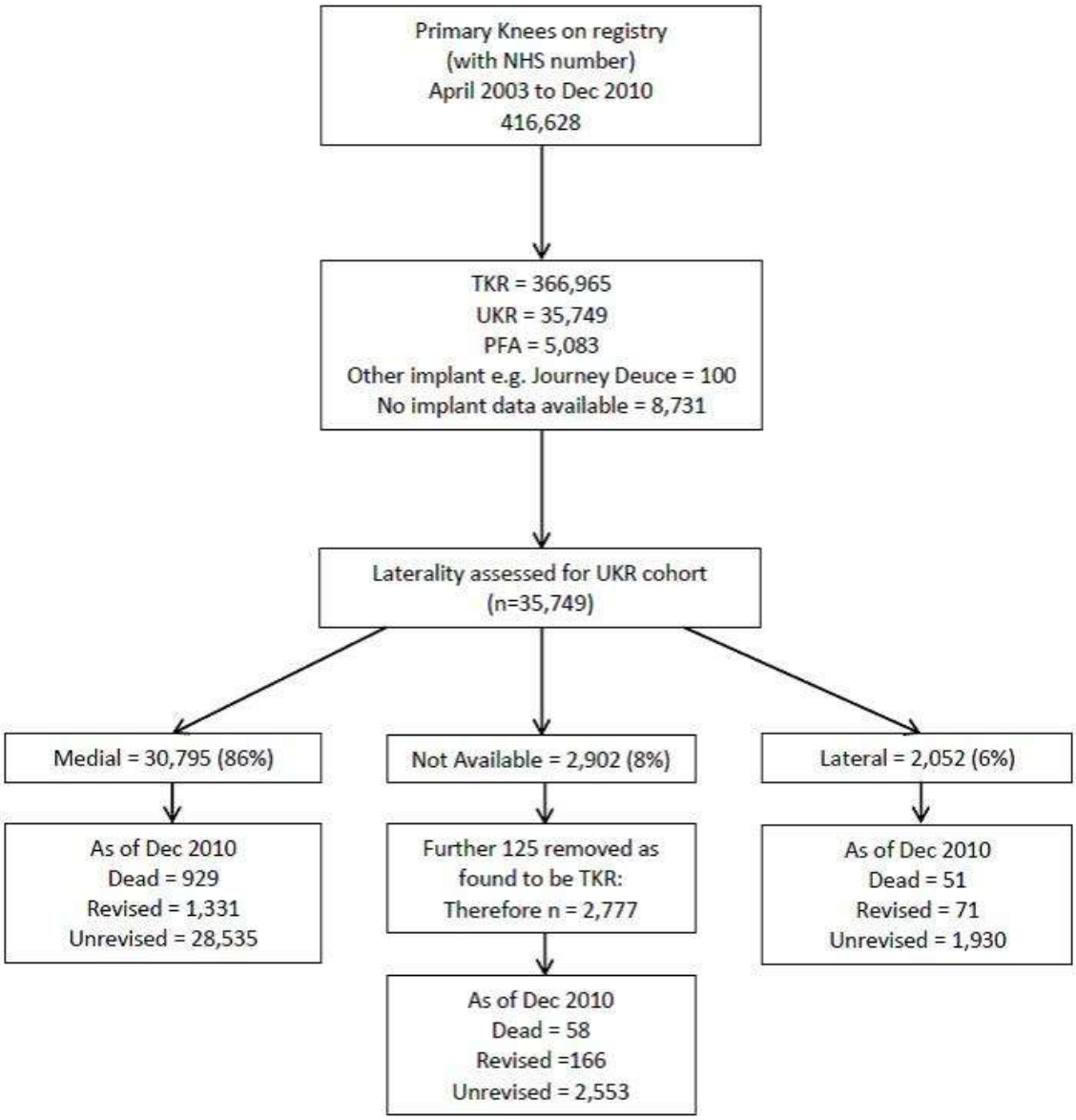


Figure 4.4: Flowchart demonstrating the preparation of the unicondylar knee replacement (UKR) dataset.

Analysis 3: Results

In total, 2052 of the 32,847 (6.2%) UKRs for which laterality could be determined were implanted on the lateral side of the knee (figure 4.4). The demographic and implant characteristics of the medial and lateral UKR groups are given in Table 4.8. The use of fixed-bearing implants was significantly higher in the lateral compartment (590 of 2052 UKRs; 29%) than in the medial compartment (5611 of 30 795 UKRs; 18%) (p<0.001).

Patient characteristics	**Medial UKR (n = 30,795)**	**Lateral UKR (n = 2052)**
Mean age (yrs) (range)	64.6 (24 to 96)	63.1 (19 to 92)
Male (n, %)	16,223 (*52.7*)	900 (*43.9*)
ASA grade (n, %)		
1	7757 (*25.2*)	614 (*29.9*)
2	20,574 (*66.8*)	1302 (*63.5*)
3 and 4	2464 (*8.0*)	136 (*6.6*)
Indication (n, %)		
Osteoarthritis	30,485 (*99.0*)	2025 (*98.7*)
Other	310 (*1.0*)	237 (*1.3*)
Implant design (n, %)		
Fixed bearing/all PE	3127 (*10.2*)	248 (*12.1*)
Fixed bearing/metal-backed	77 (*0.3*)	156 (*7.6*)
Fixed bearing/modular	2407 (*7.8*)	186 (*9.1*)
Mobile bearing/modular	24,237 (*78.7*)	1367 (*66.6*)
Not available	947 (*3.1*)	95 (*4.6*)

Table 4.8: Demographics of the medial and lateral unicondylar replacement (UKR) groups
Key: ASA, American Society of Anesthesiologists; PE, polyethylene.

Survival analyses:

The survival rate for the entire cohort at five years (all 35,749 UKRs) was 92.9% (95%CI 92.5 to 93.3) and at seven years was 90.6% (95%CI 89.9 to 21.3). Kaplan-Meier graphs and life tables for both the medial and the lateral groups are given in figure 4.5 and tables 4.9 and 4.10. The rates of survival at five years were 93.1% (95%CI 92.7 to 93.5) for medial and 93.0% (95%CI 91.1 to 94.9) for lateral UKRs. There was no statistically significant difference in the survival distributions (log rank chi-squared test = 0.48 (1 df); p=0.49).

Years since operation	Number at start	Number revised	Number withdrawn	Number at risk	Annual failure rate (%)	Survival rate (%, 95% CI)
0 to 1	30,795	320	4957	23,316.5	1.1	98.9 (98.7 to 99.0)
1 to 2	25,518	421	5335	22,850.5	1.8	97.0 (96.8 to 97.3)
2 to 3	19,762	262	5370	17,077.0	1.5	95.6 (95.3 to 95.8)
3 to 4	14,130	157	4776	11,742.0	1.3	94.3 (93.9 to 94.6)
4 to 5	9197	92	3656	7369.0	1.2	93.1 (92.7 to 93.5)
5 to 6	5449	54	2760	4069.0	1.3	91.9 (91.3 to 92.4)
6 to 7	2635	22	1812	1729.0	1.3	90.7 (90.0 to 91.4)

Table 4.9: Life table analysis for the medial unicondylar replacement group. Three implants were revised after the end of year 7 (CI, confidence interval).

Years since operation	Number at start	Number revised	Number withdrawn	Number at risk	Annual failure rate (%)	Survival rate (%, 95% CI)
0 to 1	2052	25	448	1828.0	1.4	98.6 (98.1 to 99.2)
1 to 2	1579	21	410	1374.0	1.5	97.1 (96.3 to 98.0)
2 to 3	1148	8	388	954.0	0.8	96.3 (95.3 to 97.3)
3 to 4	752	10	265	619.5	1.6	94.8 (93.4 to 96.1)
4 to 5	477	7	202	376.0	1.9	93.0 (91.1 to 94.9)
5 to 6	268	0	149	193.5	0.0	93.0 (91.1 to 94.9)
6 to 7	119	0	88	75.0	0.0	93.0 (91.1 to 94.9)

Table 4.10: Life table analysis for the lateral unicondylar replacement group (CI, confidence interval).

In order to account for differences in patient and implant covariates between the two groups, the influence of laterality was further assessed using Cox's proportional hazards. In a model including all covariates the risk of failure for medial and lateral UKRs was equivalent (hazard ratio (HR) for lateral relative to medial UKR=0.88 (95 CI 0.69 to 1.13); p=0.32).

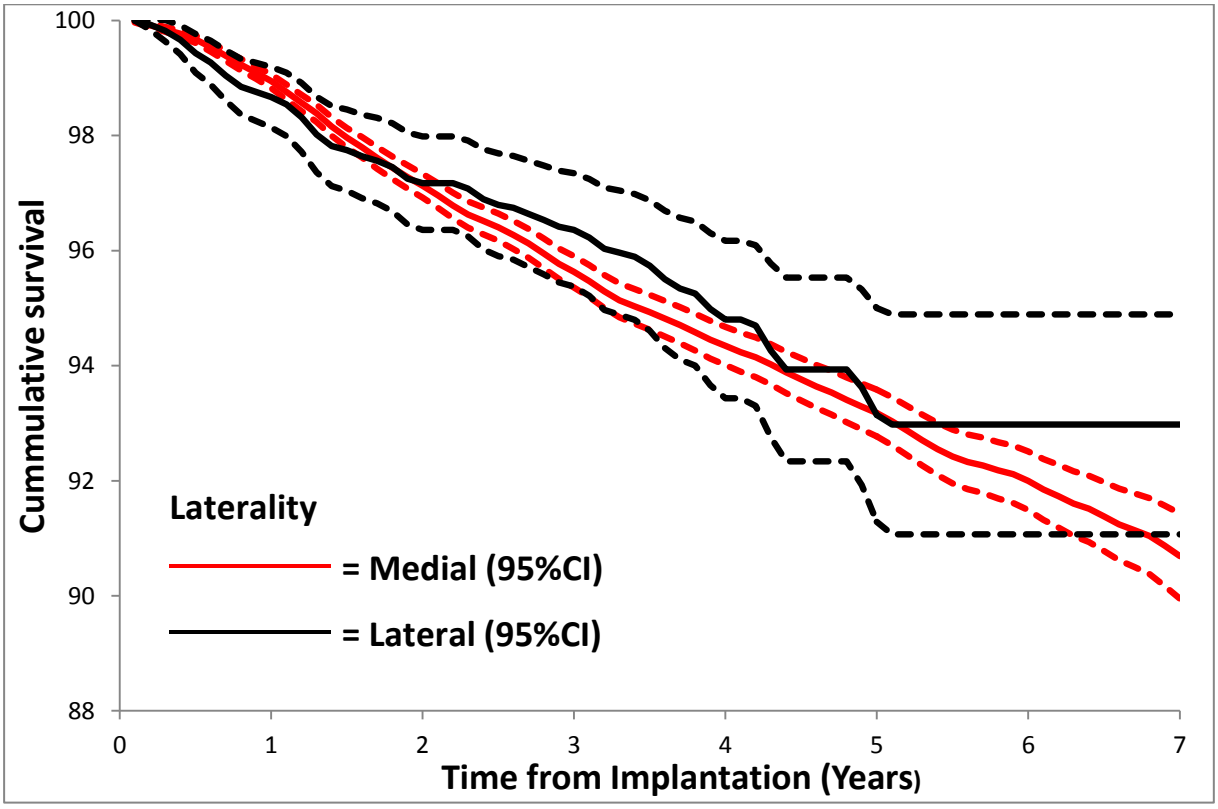


Figure 4.5: Kaplan-Meier survival curve for medial and lateral unicondylar knee replacements.

The factors influencing the risk of failure:

The full Cox models, including the influences of all the covariates on the risk of failure for both medial and lateral UKRs, are given in table 4.11. The only covariates found to influence the risk of failure for medial UKRs were patient age ($p < 0.001$) and ASA grade (p=0.03). The risk of failure decreased with increasing age (figure 4.6). Patients categorised as ASA grade 3 or 4 had a greater risk of failure than patients graded ASA 1 (HR 1.30 (95%CI 1.05 to 1.60);

p=0.02) and ASA 2 (HR 1.31 (95%CI 1.08 to 1.59); p = 0.01). Analysis of the effects of bearing on the risk of failure demonstrated no differences between mobile- and fixed-bearing implants (HR 1.05 (95%CI 0.91 to 1.20); p=0.51) (figure 4.7). There were also no differences in the risk of failure dependent upon tibial design for the fixed-bearing implants (p=0.52).

Following a similar process for the lateral UKRs, only age (p=0.01) had a significant effect in the final model. Similarly to the medial group, the risk of failure decreased with increasing age (figure 4.8). For lateral UKRs there was again no difference in the risk of failure between mobile- and fixed-bearing implants (HR 0.86 (95%CI 0.52 to 1.44); p=0.57) (figure 4.9) and the different fixed-bearing tibial designs (p=0.96).

	Medial UKR		Lateral UKR	
Covariate*	HR (95% CI)	p-value	HR (95% CI)	p-value
Age group				
< 55 years	Reference	-	Reference	-
55 to 65 years	0.75 (0.65 to 0.86)	< 0.001	0.57 (0.33 to 1.00)	0.05
65 to 75 years	0.57 (0.49 to 0.67)	< 0.001	0.36 (0.17 to 0.73)	0.005
> 75 years	0.42 (0.34 to 0.52)	< 0.001	0.35 (0.14 to 0.85)	0.02
Gender				
Female	Reference	-	Reference	-
Male	0.90 (0.81 to 1.00)	0.06	0.67 (0.41 to 1.10)	0.12
ASA grade				
1	Reference	-	Reference	-
2	0.99 (0.87 to 1.12)	0.89	0.62 (0.38 to 1.02)	0.06
3 and 4	1.30 (1.05 to 1.60)	0.02	0.76 (0.23 to 2.51)	0.65
Indication				
Osteoarthritis	Reference		Reference	-
Other	0.83 (0.44 to 1.54)	0.54	NA†	-†
Implant design				
Fixed bearing/all PE	0.99 (0.84 to 1.17)	0.91	0.96 (0.50 to 1.82)	0.89
Fixed bearing/metal-backed	1.46 (0.61 to 3.51)	0.40	0.76 (0.30 to 1.93)	0.57
Fixed bearing/modular	1.14 (0.93 to 1.42)	0.21	0.78 (0.31 to 1.97)	0.60
Mobile bearing/modular	Reference	-	Reference	-

Table 4.11: Predictors of medial and lateral unicondylar replacement (UKR) revision included in the Cox's proportional regression model (HR, hazard ratio; CI, confidence interval). Key: ASA, American Society of Anesthesiologists; PE, polyethylene, † there were no revisions in this group, so HR and p value could not be calculated.

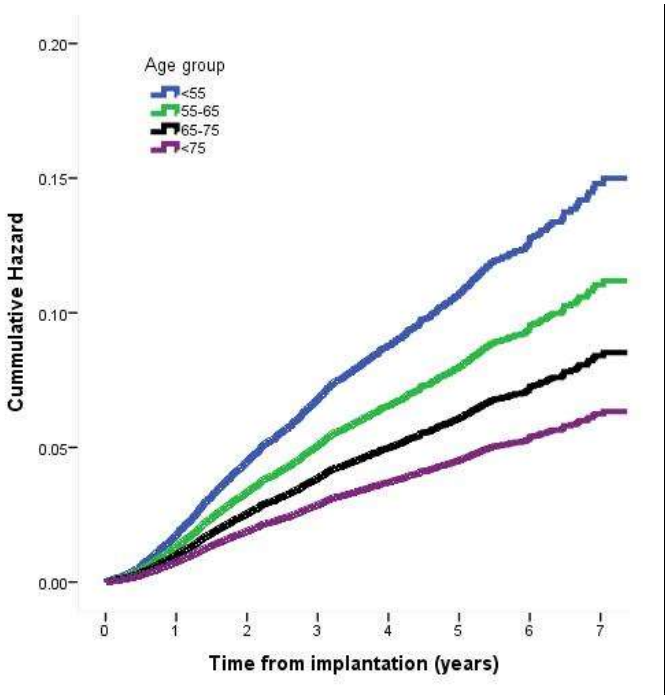


Figure 4.6: Hazard plot for the risk of failure for medial unicondylar knee replacements dependent upon patient age group after adjustment for significant covariates in the final Cox's proportional hazard model.

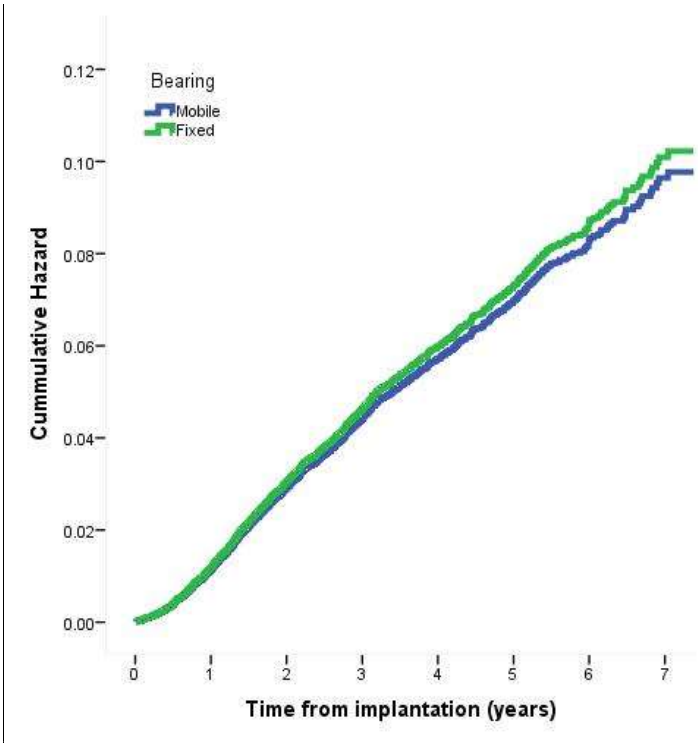


Figure 4.7: Hazard plot for the risk of failure for medial unicondylar knee replacements dependent upon implant bearing (fixed *versus* mobile) after adjustment for significant covariates in the final Cox's proportional hazard model.

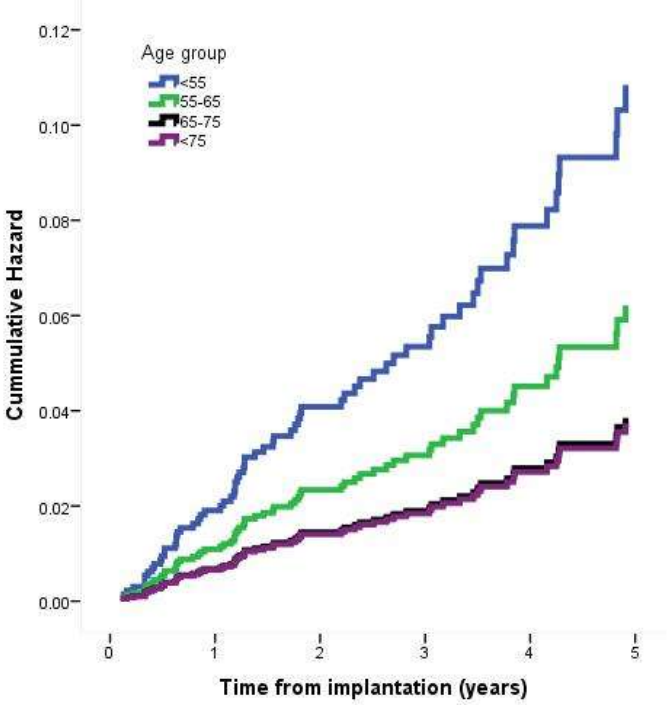


Figure 4.8: Hazard plot for the risk of failure for lateral unicondylar knee replacements dependent upon patient age group after adjustment for significant covariates in the final Cox's proportional hazard model.

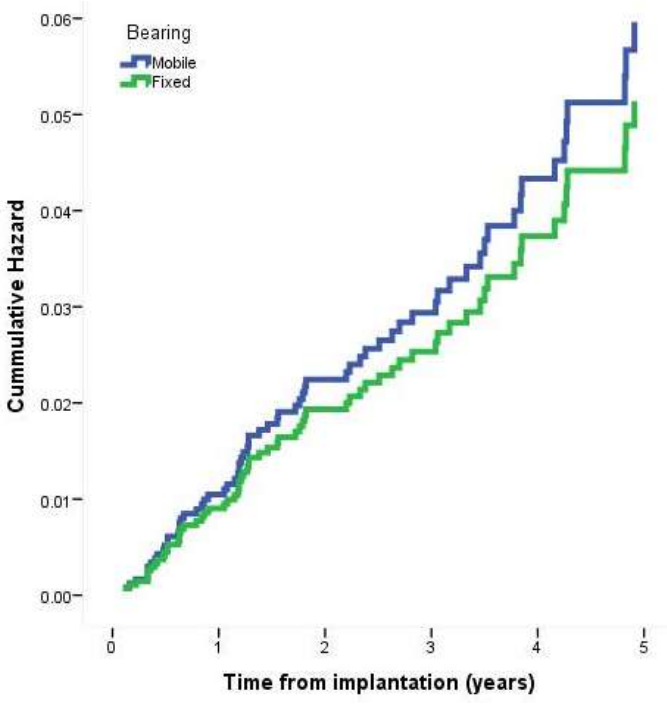


Figure 4.9: Hazard plot for the risk of failure for lateral unicondylar knee replacements dependent upon implant bearing (fixed *versus* mobile) after adjustment for significant covariates in the final Cox's proportional hazard model.

Reason for revision:

The reasons for revision for the 71 lateral and 1331 medial UKRs are given in table 4.12. The distributions of the reasons were similar for the two groups, the commonest reasons being aseptic loosening/lysis (417 of 1402 revisions, 29.7%) and unexplained pain (342 of 1402 revisions, 24.4%). Of interest was the finding that the patients revised for unexplained pain were younger (mean age 60.9 years (SD 9.2)) than those revised for all other reasons (mean age 62.3 years (SD 9.8)) (p=0.02). Although the hazard of revision for mobile- and fixed-bearing implants was similar for both lateral and medial UKRs, there were noticeable differences in the reasons for which those revisions were performed. On both sides of the knee mobile bearings were associated with a significantly greater proportion of failures for component dissociation/disability/instability than were fixed bearings (medial: mobile 137 of 1046 (13%) revisions *versus* fixed 12 of 272 (4%), p<0.001; and lateral: mobile 14 of 46 (30%) revisions *versus* fixed two of 23 (9%), p=0.04). However, this discrepancy was offset by higher rates of loosening/lysis and unexplained pain for the medial fixed-bearing UKRs, and progression of arthritis and unexplained pain with the lateral fixed-bearing UKRs.

		Medial UKR				Lateral UKR			
		Mobile	**Fixed**	**Not known**	**Total**	**Mobile**	**Fixed**	**Not known**	**Total**
Hierarchy	**Reason for revision**	**n=24,237**	**n=5611**	**n=947**	**n =30,795**	**n=1367**	**n=590**	**n=95**	**n=2052**
1	Tumour	-	-	-	-	-	-	-	-
2	Infection	53 (*5*)	19 (*7*)	1 (*8*)	73 (*5*)	4 (*9*)	3 (*13*)	-	7 (*10*)
3	Incorrect side/sizing of component	-	-	-	-	-	-	-	-
4	Malalignment	58 (*6*)	15 (*6*)	1 (*8*)	74 (*6*)	1 (*2*)	-	-	1 (*1*)
5	Metal sensitivity	-	-	-	-	-	-	-	-
6	Loosening/lysis	300 (*29*)	97 (*36*)	3 (*23*)	400 (*30*)	12 (*26*)	4 (*17*)	1 (*50*)	17 (*24*)
7	Component dissociation	33 (*3*)	2 (*1*)	-	35 (*3*)	5 (*11*)	-	-	5 (*7*)
8	Component/polyethylene wear	20 (*2*)	5 (*2*)	-	25 (*2*)	-	-	-	-
9	Implant breakage	2 (*0*)	-	-	2 (*0*)	-	-	-	-
10	Dislocation/instability	104 (*10*)	10 (*4*)	2 (*15*)	116 (*9*)	9 (*20*)	2 (*9*)	-	11 (*15*)
11	Fracture of bone	23 (*2*)	6 (*2*)	-	29 (*2*)	-	2 (*9*)	-	2 (*3*)
12	Progression of arthritis/disease	47 (*4*)	12 (*4*)	1 (*8*)	60 (*5*)	-	3 (*13*)	-	3 (*4*)
13	Synovitis	-	-	-	-	-	-	-	-
14	Arthrofibrosis/stiffness	14 (*1*)	5 (*2*)	-	19 (*1*)	-	1 (*4*)	-	1 (*1*)
15	Osteonecrosis/avascular necrosis	-	-	-	-	-	-	-	-
16	Heterotopic bone	-	-	-	-	-	-	-	-
17	Unexplained pain	247 (*24*)	75 (*28*)	5 (*38*)	327 (*25*)	8 (*17*)	7 (*30*)	-	15 (*21*)
18	Other (if not listed above)	145 (*14*)	26 (*10*)	-	171 (*13*)	7 (*15*)	1 (*4*)	1 (*50*)	9 (*13*)
	TOTAL REVISIONS	1046	272	13	1331	46	23	2	71

Table 4.12: Reasons for revision for the 1331 medial and 71 lateral unicondylar knee replacement (UKR) revisions dependent upon bearing type. Figures in brackets represent the column percentages (%) for each implant type.

Analysis 3: Discussion

This analysis demonstrates that approximately 6% of UKRs are implanted on the lateral side of the knee, and that the mid-term survival rates of lateral and medial UKRs are equivalent. Medial and lateral UKRs have a similar pattern of failure, with aseptic loosening/lysis and unexplained pain the predominant reasons for revision. The main factor influencing the risk of failure was patient age, and this finding was consistent for both medial and lateral UKRs.

Approximately one in 15 UKRs were implanted laterally, slightly less than the one in ten suggested by Scott (Scott 2005). Owing to the failure to report medial and lateral UKRs separately, it is difficult to obtain an idea of lateral UKR usage from other registry reports. No information is available from the current Australian (AJR-AR 2010) New Zealand (NZJR-AR 2010) and Norwegian (Norwegian registry-AR 2010) registries. Data from the Swedish knee register suggest that their usage is significantly lower (three of 683 UKRs (0.4%)) (Lidgren 2009) than was seen in this analysis, and has declined substantially from the 1336 (9%) of 14 772 UKRs reported by the same registry in the mid-1990s (Lewold 1998).

The overall seven-year survival rate for all UKRs, including those for which laterality could and could not be determined, was 90.6%. This is similar to the 88.9% seven-year survival rate reported by the Australian registry (AJR-AR 2010) but significantly lower than the 97.3% at seven years reported by the originating centre for the Oxford UKR (Pandit 2011). This highlights the discrepancy in the literature between outcomes from registries, independent studies and inventor studies for UKR. In a recent literature review Labek *et al* (Labek 2011) found that, compared with studies from the originating centre, the reported rates of failure were 2.5 times higher in independent clinical studies and over four times higher in national arthroplasty registers. At five years, the rates of survival for medial and lateral UKRs were equivalent. This remained the case at seven years, although there were no further failures of lateral UKRs between years five and seven. The rates of survival for lateral UKRs were lower than those seen in other contemporary series, in which survival rates of 92% (38

UKRs) and 98% (54 UKRs) at ten years (Ashraf 2002, Argenson 2008) and 100% at a mean of 5.2 years (49 UKRs) (Sah 2007) and 12.4 years (28 UKRs) (Pennington 2006) have been reported.

For both medial and lateral UKRs the key determinant of failure was the patient's age at primary surgery. A decline in revision rate with increasing age had been reported by a number of registries across a range of types of UKR (Lidgren 2007, AJR-AR 2010). For medial UKR the risk of failure was greater for patients graded ASA 3 or 4 than for those with ASA grades 1 and 2. This is likely to be related to a combination of factors, including an elderly population which puts less demand on their replacement and an increasing unwillingness of surgeons and patients to commit to revision surgery for the failing implant as age and comorbidity increase (Goodfellow 2010). Although the groups were well matched for patient demographics, there were differences in the proportions of fixed and mobile bearings used in the medial and lateral compartments, with more fixed bearings being implanted laterally (29% *versus* 18% medially). Mobile-bearing UKRs have better kinematics than fixed-bearing UKRs (Li 2006), with different patterns of wear and modes of failure (Kretzer 2011). However, this study found that the failure of both medial and lateral UKRs at seven years was not influenced by the type of bearing or, for fixed-bearing implants, the design of the tibial component. A number of registries have reported variations in UKR revision rates depending on implant brand (AJR-AR 2010, NZJR-AR 2010, NJR-AR 2012). Our analysis did not examine the effect of brand, as brand and design are not independent of one another and therefore cannot be examined together. Nevertheless, it is likely that any analysis of brand would have mirrored the results found in the NJR (NJR-AR 2012).

The distribution of the reasons for revision was similar for both medial and lateral UKRs, with aseptic loosening/lysis and unexplained pain being the main causes. Patients whose UKR was revised for unexplained pain were on average 1.4 years younger than those whose UKR was revised for any other reason. This finding is reflected in the New Zealand (NZJR-AR 2010) and Norwegian (Norwegian registry-AR 2010) registries, where aseptic loosening and

pain are the predominant reasons for revision. Analysis of the Norwegian registry has also shown that UKRs have a greater risk of revision due to aseptic loosening and pain than TKRs (Furnes 2007). Although the overall rates of revision were similar for fixed- and mobile-bearing UKRs, more revisions were performed for dissociation/dislocation/instability in the mobile-bearing group. This difference was particularly marked in the lateral UKR group, where 30% of mobile-bearing implants were revised for this reason. This may reflect the differential kinematics of the medial and lateral compartments (Freeman 2005) and emphasises the importance of correct positioning of the components and soft tissue balancing when using mobile-bearing UKRs.

Analysis 3: Summary

This analysis demonstrates that the rates of survival and reasons for revision for medial and lateral UKR are similar. There is no evidence to suggest that the performance of either type is adversely affecting the results seen for the other when the results are pooled for the purposes of registry analysis of UKRs.

Analysis 3: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

Within the NJR the majority of UKRs (>70%) are Oxford brand implants. Any generic analysis of UKR, such as undertaken here, is therefore likely to be overwhelmed by the influence of this implant type and may not be a true reflection of the behaviour of other UKRs that represent a much smaller percentage of the total UKRs performed. Because of this we wanted to perform a sub analysis in which the effects of the different types of UKR could be assessed. Unfortunately the NJR did not permit this level of analysis.

This may have compromised certain elements of the analysis, specifically the analysis of the impact of bearing (fixed *versus* mobile) upon the risk of revision. The mobile bearing group almost exclusively consisted of Oxford UKRs, whereas the fixed bearing group comprised a variety of different implant types. When these fixed bearing implants were pooled the risk of revision was equivalent to that seen with the mobile bearing implants. However, the NJR (NJR-AR 2012) reports a wide variation in the mid-term survival of the fixed bearing implants with some having very low failure rates and some having very high failure rates. By including information on implant type we could have determined which implants worked best in each compartment of the knee. Instead by pooling implants with differing performance we have lost the necessary detail to address this issue. This highlights an ongoing problem with registry data, namely that "the devil is in the detail". Broadly pooling implants into groups for convenience and simplification can miss important associations within the data that might have been evident if a greater level of detail had been employed.

The NJR does not currently record whether a UKR is implanted in the medial or lateral compartment. Consequently we had to use implant data combined with the side of surgery to determine implant laterality. Using this method we were able to determine laterality for the majority (92%) but not all of the UKR implants. A solution to this problem would be for the NJR K1 form to be amended so that the surgeon can indicate UKR laterality at the time the form is completed. This data would then be stored within the dataset, ensuring higher levels of data completion and accuracy and avoiding the need to combine data from a number of different data fields.

Unfortunately we were unable to perform a meaningful comparison of functional outcomes for the lateral and medial UKRs using the NJR-PROMs dataset. Having determined the laterality of each UKR in the NJR dataset we were then able to determine the laterality of the UKR in the NJR-PROMs dataset. However, of the 505 UKR within the PROMs dataset only 21 were lateral UKR and the small numbers of lateral implants precluded any meaningful comparison with the medial UKR group. As the PROMs project matures and

more PROMs become available it is likely that the number of PROMs records linked to medial and lateral UKR will increase. This issue could therefore be re-visited in the future, allowing the function of medial and lateral UKR to be directly compared.

The identification of the lateral UKR group opens up the possibility of further work into this group of patients. We could identify which surgeons are undertaking lateral UKRs, allowing us to better understand when and why they are being performed. It could allow us to determine how many procedures the average surgeon undertakes each year. By directly contacting these surgeons we could determine whether they employ different selection criteria when selecting patients and employ different surgical processes for their medial and lateral UKR patients.

4.5 Analysis 4: Analysis of the effect of centre and surgeon operative volume on UKR rates of failure

Analysis 4: Aim:

To determine:

1. How many different centres and surgeons are performing UKR in England and Wales and what volumes of these procedures are they undertaking.
2. Is there an association between centre and/or surgeon volume and revision rate for UKR.
3. If centres/surgeons perform high volumes of UKR do they achieve revision rates comparable to those seen with the 'gold standard' cemented TKR.

Analysis 4: Methods

The 35,749 UKR recorded in NJR dataset described in section 3.4.2 were again used as the basis for analysis. The rates of UKR failure vary dependent upon implant type (NJR-AR 2012, AJR-AR 2010) and as such this variable could be a source of confounding in any comparison of centre and surgeon revision rates. A decision was therefore made to limit the analysis to the Oxford UKR (Biomet) as it is the most commonly registered UKR within the NJR comprising (n=26,293) 74% of all available implants. From this group we selected all cemented UKRs implanted in the medial compartment for osteoarthritis. This was done because, despite the previous analysis suggesting that when all implants were pooled the overall rates of failure for medial and lateral UKR were similar, we could not be certain that this was the case for the Oxford UKR. The method for determining implant laterality (medial *versus* lateral) is discussed in section 4.4. The additional data relating the type of fixation (cemented/uncemented/hybrid) and indications for surgery was available within the NJR dataset. Details of the numbers of UKR excluded from the initial 35,749 cohort are given in figure 4.10. In total 23,400 Oxford UKRs satisfied these criteria.

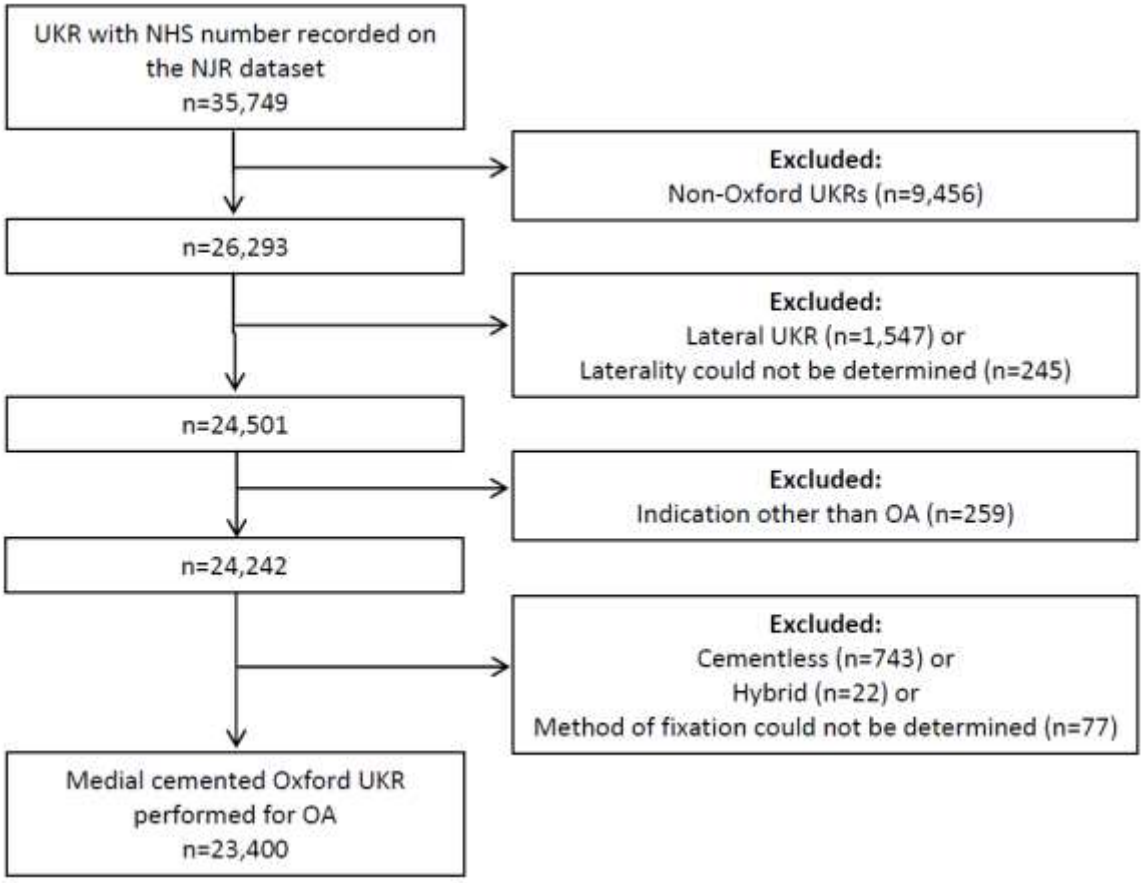


Figure 4.10: Relevant exclusions for Analysis 4.

The effect of operative volume was analysed by considering volume as both a continuous and categorical variable. For the categorical analysis both centre and surgeon volumes were grouped based upon the volume of primary procedures performed during the period of observation. Centres were grouped to produce 5 categories containing approximately equal number of primary procedures while maintaining rounded category boundaries (1-50, 51-100, 101-200, 201-400, >400). Surgeons were similarly grouped into 5 approximately sized groups (1-25, 26-50, 51-100, 101-200, >200). As these groups contained larger numbers it was then possible to construct Kaplan Meier plots and life tables to supplement the funnel plots examining the number of revisions per 100 component years for each individual centre and surgeon.

Analysis 3 (section 4.4) demonstrated that patient factors including age and ASA grade independently influence UKR failure rates. Adjustments were therefore made for differences in these factors between the volume groups using Cox's proportional hazards. Independence of centre and surgeon volume cannot be assumed and it is therefore inappropriate to include them in the same hazard model. Initial analyses were therefore performed with centre and surgeon volume modeled separately using hazard analyses that also included age group (<55/55-64.9/65-74.9/>75 years), gender (Male/Female) and ASA grade (ASA 1/2/3 and 4). As we were interested in whether higher volumes were associated with improved outcomes we chose the highest centre/surgeon volume category as the reference within each of the hazard models. To supplement the analysis of the grouped volume data we constructed additional models in which centre and surgeon volume were entered as continuous explanatory variables based on their total volume over the period of study. Other than entering volume data as a continuous rather than a categorical variable these models were identical.

The individual centre and surgeon volume analyses were then supplemented by a further analysis that considered these variables simultaneously. Centres and surgeons were categorised as low volume or high volume based on volumes of ≤100/>100 as determined by the initial analyses. This gave four groups (low centre volume/low surgeon volume, low centre volume/high surgeon volume, high centre volume/low surgeon volume, high centre volume/high surgeon volume) which were then modelled using Cox's proportional hazards adjusting for age, gender and ASA grade.

Analysis 4: Results

The demographic characteristics of study cohort are given in table 4.13. Even with the exclusions imposed the study cohort was a well matched, representative sample of the wider population undergoing UKR. For the entire study cohort the 5-year and 7-year survival rates were 93.2% (95%CI 92.8 to 93.7%) and 90.7% (95%CI 89.9 to 91.5%)

respectively. The overall rate of revisions per 100 component years was 1.40 (95%CI 1.31 to 1.49).

Demographic variable	Study Cohort (Medial cemented Oxford UKR) (April 2003 to December 2010) n=23,400	NJR 8th Annual report (All UKR) (2010 data only) n=6,119
Gender		Gender data available for 5,702
Female	11,207 (48%)	2,675 (47%)
Male	12,193 (52%)	3,027 (53%)
ASA grade		
ASA 1	5,797 (25%)	1,421 (23%)
ASA 2	15,681 (67%)	4,221 (69%)
ASA 3/4	1,922 (8%)	477 (8%)
Age groups		Age data available for 5,702
<55 years	3,379 (14%)	1072 (19%)
55-64.9 years	8,668 (37%)	2,051 (36%)
65-74.9 years	7,723 (33%)	1,800 (32%)
>75 years	3,630 (16%)	779 (14%)
Mean Age (S.D)	65.0 (9.3)	64.0 (9.8)

Table 4.13: Patient demographics for analysis 4.

Centre and surgeon volumes

In total 366 centres performed at least 1 primary procedure during the period of analysis. The median number of primary procedures was 33 with a range of 1 to 1097 (figure 4.11). In total 303 (82.8%) performed 100 or less primary procedures. The rate of revisions per 100 component years varied widely between institutions ranging from 0.00 to 41.49. The variability in this measure and the number of centres above 3 standard errors of the mean value increased as centre volume decreased (figure 4.12). In total 46 of the 303 centres (15.2%) that performed ≤100 primary procedures, 6 of the 39 centres (15.4%) that performed 100-200 and 2 of the 24 centres (8.3%) that performed >200 procedures were above 3 standard errors of the mean.

Nine-hundred and nineteen surgeons performed at least 1 primary procedure during the period of analysis. The median number of primary procedures was 8 with a range of 1 to 423 (figure 4.13). In total 787 (85.6%) performed 50 or less primary procedures. The rates of revisions per 100 component years for individual surgeons ranged from 0.00 to 833.33. In similarity to centre volume the variability in this measure and the number of centres above 3 standard errors of the mean value increased as surgeon volume decreased (figure 4.14). Of the 787 surgeons that performed ≤50 procedures 128 (16.3%) were above the 3 standard error threshold compared to 5 of 81 (6.2%) and 4 of 51 (7.8%) surgeons who performed 50-100 and >100 respectively.

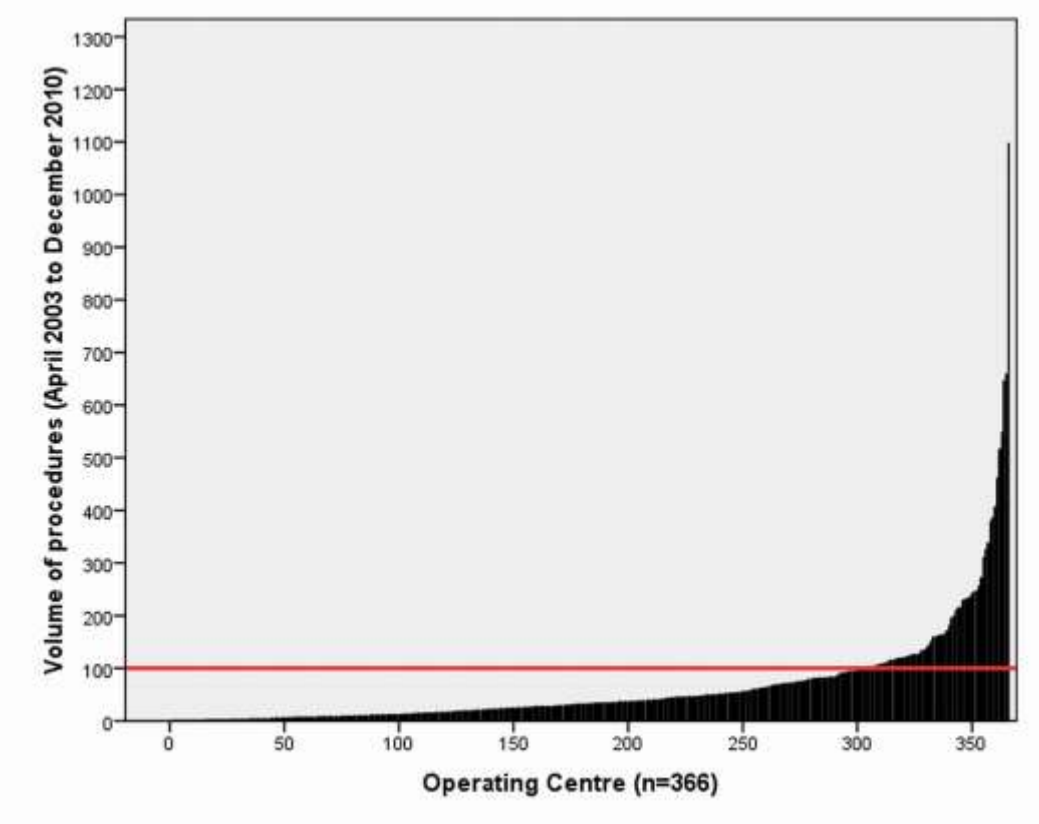


Figure 4.11: Volume for each of the 366 centres undertaking procedures between (April 2003 and December 2010).

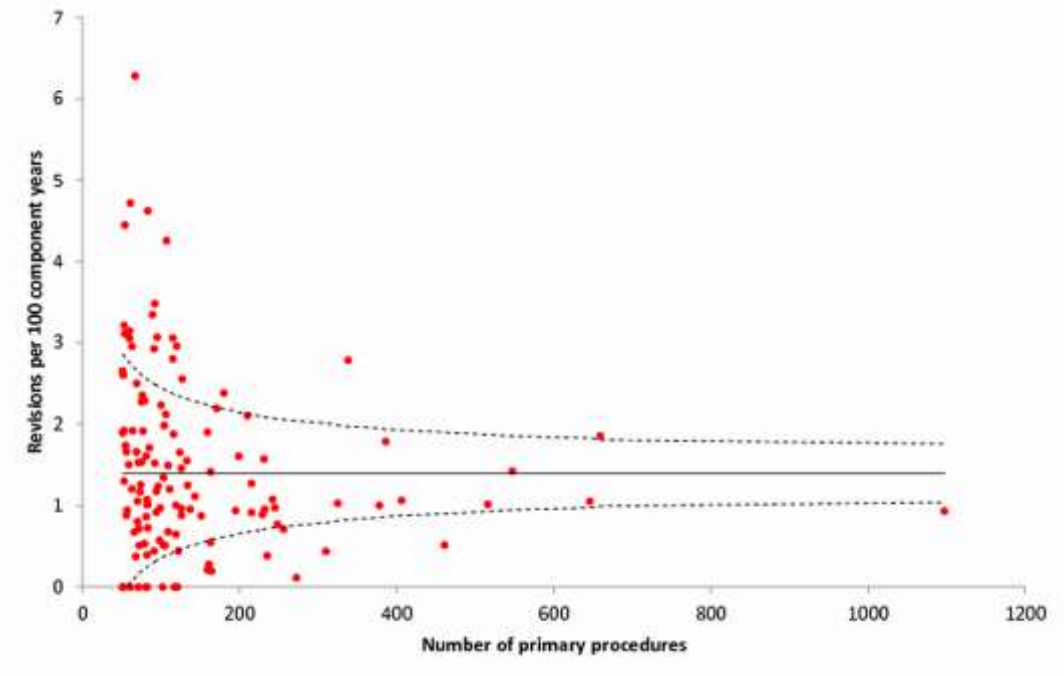


Figure 4.12: Funnel plot of revisions per 100 component years against centre volume: April 2003 to December 2010 (Dotted lines represent +/-3 standard errors from the mean) (Only centres performing ≥50 procedures included for clarity).

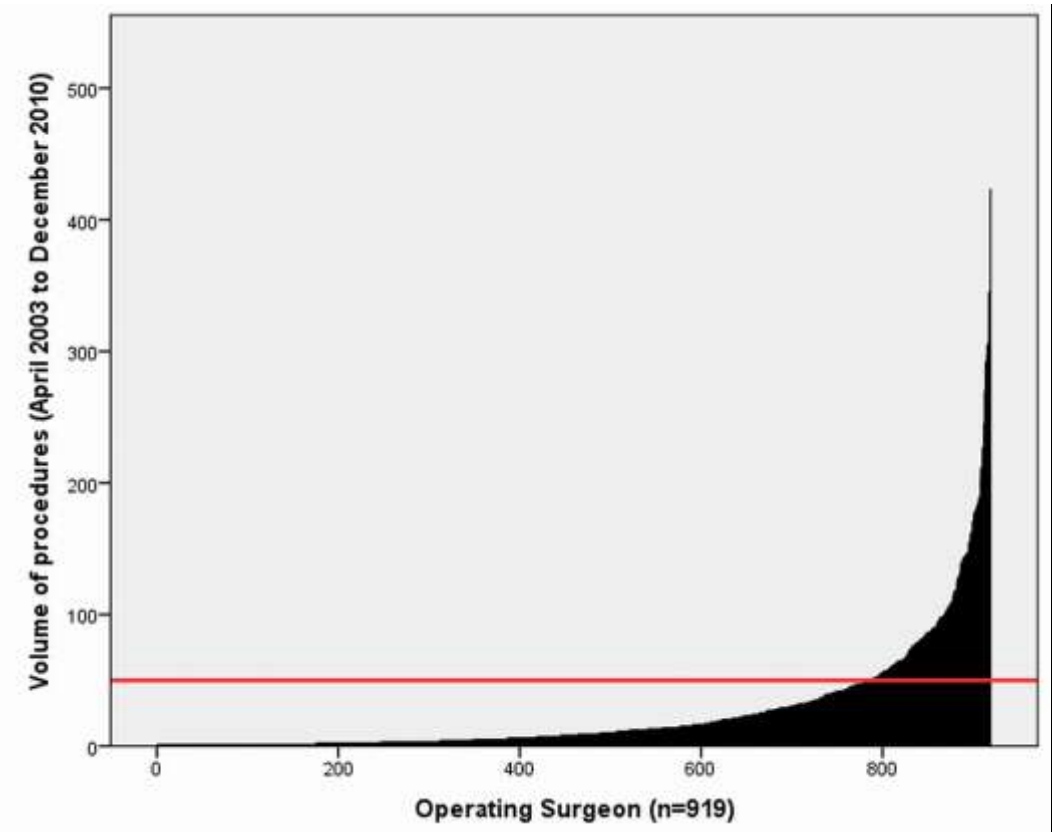


Figure 4.13: Volume for each of the 919 surgeons undertaking procedures between (April 2003 and December 2010).

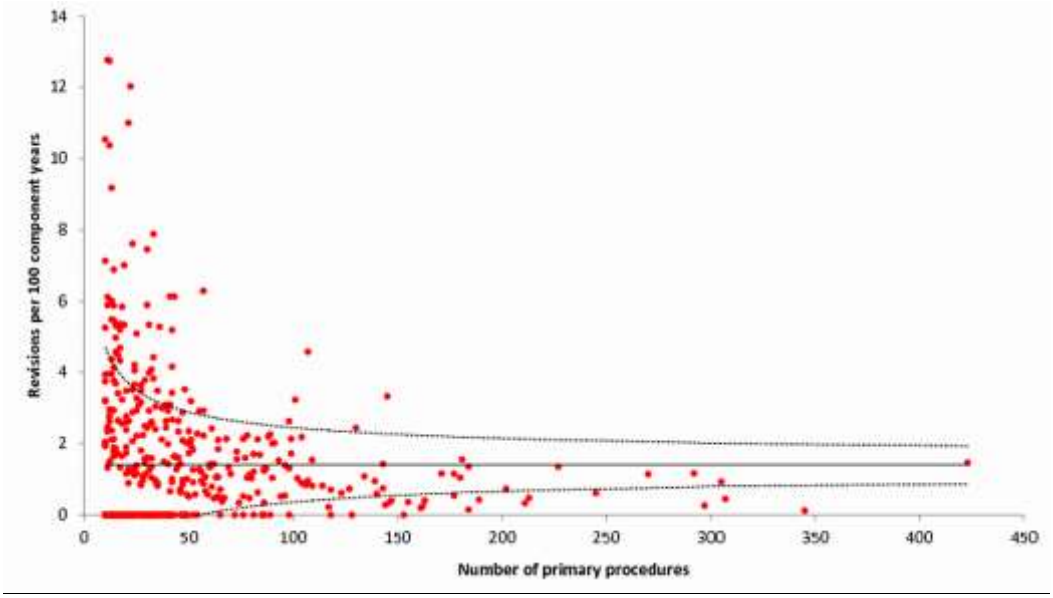


Figure 4.14: Funnel plot of revisions per 100 component years against surgeon volume: April 2003 to December 2010 (Dotted lines represent +/-3 standard errors from the mean) (Only surgeons performing ≥10 procedures included for clarity).

Volume grouping analysis

The number of revisions per 100 component years, ratio of revision:primary cases and the 5- and 7-year survival rates were calculated for each of the 5 centre (1-50, 51-100, 101-200, 201-400, >400) and surgeon (1-25, 26-50, 51-100, 101-200, >200) volume categories (tables 4.14 and 4.15). Kaplan-Meier survival plots for each analysis are shown in figures 4.15 and 4.16. Comparison of failure rates for the 5 groups confirmed that there were significant differences in the rates of failure between one or more groups ($p<0.001$). Kaplan Meier analysis (figure 4.15) demonstrated that the rates of failure were similar for the lowest two volume categories (<50 and 51-100) and that the rates for these two groups were higher than the rates of failure for the highest two volume categories (201-400, >400). The early failure rate in the intermediate volume group (101-200) mirrored that of the lowest volume groups before "crossing over" to the highest volume groups between years 2 to 4. After year 4 the failure rates for this group was similar to that seen for the two highest volume groups.

In similarity to the centre volumes there were significant differences between one or more of the surgeon volume groups ($p<0.001$). However, in contrast to the survival plots for centre volume the volume groups within the surgeon volume analysis were not clustered. They instead showed an additive pattern with each additional increase in volume associated with a better rate of survival when compared to the previous volume group (figure 4.16).

Volume of primaries (Apr 03 to Dec 10)	Number of centres	No. of primary cases	No. of revision cases	Revisions/ primaries (%)	Observed component years	Revisions per 100 observed component years	Range	5 year survival	7 year survival
1 to 50	239	4689	245	5.2%	15076.9	1.62 (95%CI 1.42 to 1.82)	0.00 to 41.49	92.3% (95%CI 91.2 to 93.3%)	89.5% (95%CI 87.8 to 91.1%)
51 to 100	64	4648	242	5.2%	14207.4	1.70 (95%CI 1.49 to 1.92)	0.00 to 6.28	92.0% (95%CI 90.9 to 93.1%)	89.2% (95%CI 87.3 to 91.1%)
101 to 200	39	5164	210	4.1%	16213.1	1.30 (95%CI 1.12 to 1.47)	0.00 to 4.25	94.2% (95%CI 93.4 to 95.1%)	92.3% (95%CI 90.8 to 93.6%)
201 to 400	17	4567	164	3.6%	13862.4	1.18 (95%CI 1.00 to 1.36)	0.11 to 2.79	93.5% (95%CI 92.4 to 94.6%)	90.8% (95%CI 88.6 to 93.0%)
>400	7	4332	139	3.2%	11943.1	1.16 (95%CI 0.97 to 1.36)	0.51 to 1.86	94.1% (95%CI 93.0 to 95.2%)	91.5% (95%CI 89.0 to 94.0%)
Total	**366**	**23400**	**1000**	**4.3%**	**71302.9**	**1.40 (95%CI 1.31 to 1.49)**	**0.00 to 41.49**	**93.2% (95%CI 92.8 to 93.7%)**	**90.7% (95%CI 89.9 to 91.5%)**

Table 4.14: Revision outcome grouped by centre volume.

Volume of primaries (Apr 03 to Dec 10)	Number of surgeons	No. of primary cases	No. of revision cases	Revisions/ primaries (%)	Observed component years	Revisions per 100 observed component years	Range	5 year survival	7 year survival
1 to 25	671	4511	287	6.4%	13296.2	2.16 (95%CI 1.19 to 2.41)	0.00 to 833.33	90.1% (95%CI 88.8 to 91.3%)	87.1% (95% CI 85.1 to 89.0%)
26 to 50	116	4269	234	5.5%	13310.0	1.75 (95%CI 1.53 to 1.98)	0.00 to 7.89	91.7% (95%CI 90.5 to 92.8%)	88.5% (95%CI 86.4 to 90.6%)
51 to 100	81	5841	225	3.9%	17469.8	1.29 (95%CI 1.12 to 1.46)	0.00 to 6.29	93.4% (95%CI 92.4 to 94.4%)	91.3 (95%CI 89.8 to 92.9%)
101 to 200	39	5442	177	3.3%	17596.7	1.01 (95%CI 0.86 to 1.15)	0.00 to 4.58	95.2% (95%CI 94.4 to 95.9%)	92.9% (95%CI 91.3 to 94.4%)
>200	12	3337	77	2.3%	9630.2	0.80 (95%CI 0.62 to 0.98)	0.12 to 1.46	96.0% (95%CI 95.0 to 97.0%)	94.3% (95%CI 92.3 to 96.2%)
Total	**919**	**23400**	**1000**	**4.3%**	**71302.9**	**1.40 (95%CI 1.31 to 1.49)**	**0.00 to 833.33**	**93.2% (95%CI 92.8 to 93.7%)**	**90.7% (95%CI 89.9 to 91.5%)**

Table 4.15: Revision outcome grouped by surgeon volume.

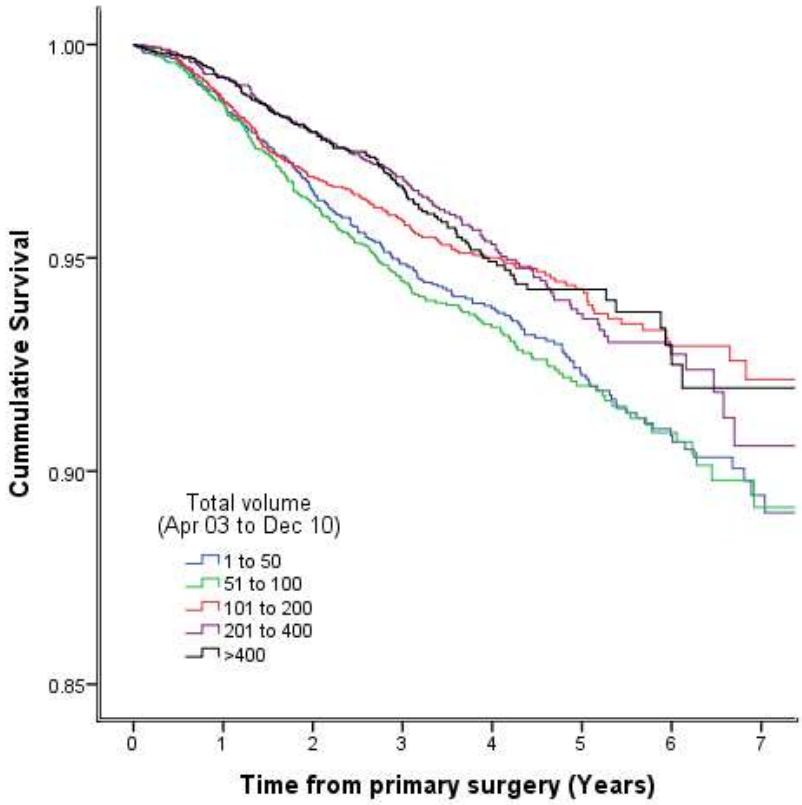


Figure 4.15: Kaplan Meier survival for UKR grouped by centre volume.

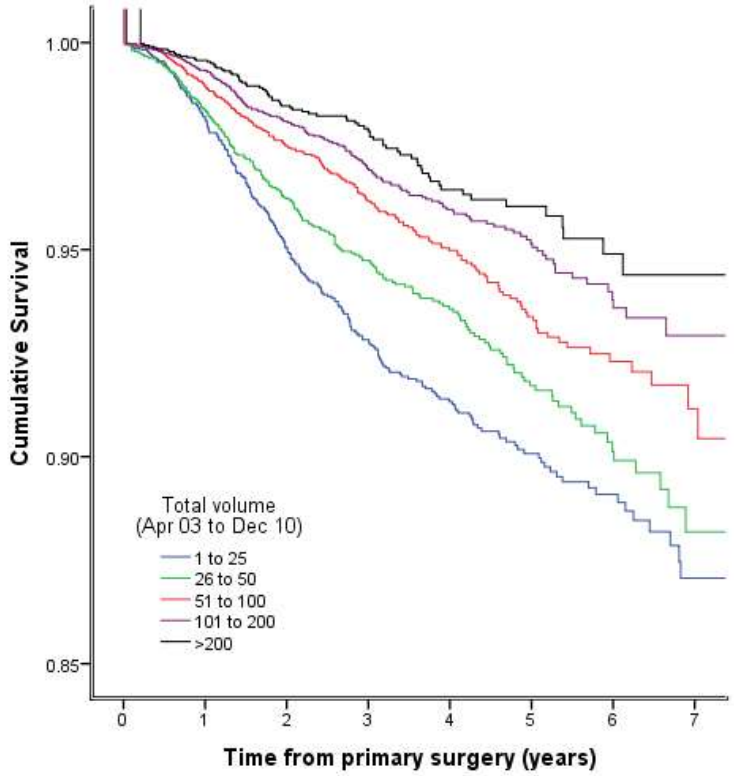


Figure 4.16: Kaplan Meier survival for UKR grouped by surgeon volume.

Hazard analysis

The models for which the volume measures were entered as continuous variables demonstrated that as centre (Wald = 8.479 (1df), HR 1.000 (95%CI 0.999 to 1.000), p=0.004) and surgeon (Wald = 68.231 (1df), HR 0.996 (95%CI 0.996 to 0.997), p<0.001) volumes increased the risk of revision decreased. The observation that the Wald test statistic was comparatively larger for surgeon volume in comparison to centre volume in identically constructed models suggests surgeon volume is the more important of these two variables.

The analysis of the grouped centre volume data demonstrated that the risk of failure was higher in the lowest (1-50, p=0.005) and second lowest (51-100, p=0.001) volume groups when compared to the highest (reference) volume group (table 4.16). The rates of failure in these two lowest volume groups were also higher than those observed for the 101-200 (HR vs. 1-50 = 1.23 (95%CI 1.02 to 1.48), p=0.03; HR vs. 51-100 = 1.30 (95%CI 1.08 to 1.56), p=0.006) and 201-400 (HR vs. 1-50 = 1.30 (95%CI 1.07 to 1.59), p=0.009; HR vs. 51-100 = 1.38 (95%CI 1.13 to 1.67), p=0.002) groups. There were no differences between the three highest volume groups within this model.

	Unadjusted (Univariate)			Adjusted (Multivariate)*		
	Hazard ratio	95.0% CI for HR	p value	Hazard ratio	95.0% CI for HR	p value
Total volume						
-1 to 50	1.40	1.14 to 1.73	0.001	1.35	1.10 to 1.66	0.005
-51 to 100	1.47	1.19 to 1.81	<0.001	1.42	1.16 to 1.76	0.001
-101 to 200	1.11	0.90 to 1.38	0.33	1.10	0.88 to 1.36	0.40
-201 to 400	1.02	0.81 to 1.28	0.88	1.03	0.83 to 1.30	0.77
->400	Ref	Ref	-	Ref	Ref	-

Table 4.16: Influence of centre volume on the risk of revision (*Adjusted for Age group, ASA grade and gender).

For surgeon volume the risk of failure was significantly higher for the three lowest volume groups (1-25, 26-50, 51-100) when compared to the highest (reference) volume group (>200) (all p<0.001) (table 4.17). The rates of failure were also significantly higher for the three lowest volume groups when compared to the 101-200 group (HR vs. 1-25 = 2.04 (95%CI 1.69 to 2.47), p<0.001; HR vs. 26-50 = 1.69 (95%CI 1.39 to 2.6), p<0.001; HR vs. 51-100 = 1.26 (95%CI 1.03 to 1.53), p=0.02). There were no differences between the two highest volume groups within this model.

	Unadjusted (Univariate)			Adjusted (Multivariate)*		
	Hazard ratio	95.0% CI for HR	p value	Hazard ratio	95.0% CI for HR	p value
Total volume						
-1 to 25	2.72	2.12 to 3.50	<0.001	2.54	1.97 to 3.27	<0.001
-26 to 50	2.20	1.70 to 2.85	<0.001	2.11	1.63 to 2.72	<0.001
-51 to 100	1.61	1.24 to 2.09	<0.001	1.56	1.21 to 1.63	<0.001
-101 to 200	1.25	0.96 to 1.64	0.10	1.24	0.95 to 1.63	0.11
->200	Ref	Ref	-	Ref	Ref	-

Table 4.17: Influence of surgeon volume on the risk of revision (*Adjusted for Age group, ASA grade and gender).

The combined effect of centre and surgeon volume was assessed by categorising centres/surgeons as high (>100 primaries) or low (≤100 primaries) volume (table 4.18). In total 7,282 primaries were performed by the high centre volume/high surgeon volumes group. Compared to the rates for the entire cohort the rate of revisions per 100 component years (0.91, 95%CI 0.79 to 1.04) was lower, and the 5- (95.6%, 95%CI 94.9 to 96.3%) and 7-year (93.2%, 95%CI 91.8 to 94.7%) survival rates were higher (all p<0.001) in this “best case” group. The hazard of revision was significantly higher for the low centre volume/low surgeon volume (HR 1.87 (95%CI 1.58 to 2.22), p<0.001) and the high centre volume/low surgeon volume (HR 1.66 (95%CI 1.39 to 1.98), p<0.001) groups when compared to the “best case” high volume group. There was no difference between the low centre volume/high surgeon volume group and the “best case” high volume group (HR 1.10 (95%CI

0.81 to 1.50), p=0.55). These findings are further evidence that surgeon volume is the more important of these two variables.

Volume	Surgeon volume	
Centre volume	Low volume (n=14,621)	High volume (n=8,779)
Low volume (n=9,337)	n=7,840 5-year survival = 91.6% *(95%CI 90.7 to 94.4%)* 7-year survival = 88.6% *(95%CI 87.2 to 90.0%)*	n=1,497 5-year survival = 94.9% *(95%CI 93.3 to 96.4%)* 7-year survival = 93.6% *(95%CI 91.5 to 95.7%)*
High volume (n=14,063)	n=6,781 5-year survival = 92.2% *(95%CI 91.3% to 93.2%)* 7-year survival = 90.0% *(95%CI 88.3 to 91.6%)*	n=7,282 5-year survival = 95.6% *(95%CI 94.9 to 96.3%)* 7-year survival = 93.2% *(95%CI 91.8 to 94.7%)*

Table 4.18: 5 and 7-year survival rates when centre and surgeon volume are considered simultaneously. High volume = >100 primary UKR between April 2003 and December 2010, Low volume = ≤100 primary UKR over same period.

Analysis 4: Discussion

This analysis demonstrates there is substantial variation in the volumes of cemented medial Oxford UKRs performed by centres and surgeons in England and Wales. A large number of centres and surgeons undertook these procedures, the majority of which did so at low volumes (<100 over lifespan of the registry). While the risk of revision decreased as both centre and surgeon increased the effect was more pronounced for surgeon volume. The risk of revision was significantly greater for centres and surgeons who had performed less than

100 procedures over the period of observation when compared to those performing greater than 100. This equates to approximately 13 procedures per year during this time period. For centres and surgeons performing greater than this number the failure rates were less than 1% per year at a maximum of 7 years.

The associations between volume and outcome found in this analysis have previously been reported for centre volume (Robertsson 2001, Hamilton 2010, Labek 2011) but not for surgeon volume for which the association with revision outcome was more pronounced. Increasing volume decreased both revision rate and reoperation rate in the 445 patients reviewed by Hamiton *et al*. (Hamilton 2001). For the first half of UKR performed the 2-year revision rate was 5.0% compared to 2.7% for the second half. Reoperation rates similarly fell from 8.1% to 5.4%. As previously discussed revision rates vary dependent upon the publishing institution (Labek 2011). Our overall rate of revisions per 100 component years (1.40, 95%CI 1.31 to 1.49) was comparable to that from other registries (1.60, 95%CI 1.43-1.80). The revision rate for centres performing the highest volumes, which included the inventing centre (1.16 (95%CI 0.97 to 1.36), was significantly higher than the pooled results from independent clinical studies including "inventor" studies (0.70, 95%CI 0.60 to 0.82) reported by Labek *et al*. (Labek 2011). The results for this group were instead in keeping with the pooled results from independent clinical studies excluding "inventor" studies (1.22, 95%CI 1.99 to 1.50). This suggests disparity in the reported results from the inventing centre for this implant type.

National analysis of UKR cost-effectiveness suggests TKR may be a more cost-effective option given the high rates of revision associated with UKR (Koskinen 2008). However, the calculations on which these conclusions are based are derived from registry revision rates and not those centres and surgeons undertaking high volumes for which the results are substantially better. One hundred and thirty seven of the 919 (14.9%) surgeons undertaking UKR fell outside the 3 standard error limit for revisions per 100 component years, the majority of whom (133) performed ≤100 procedures. If surgical workload was performed by

fewer higher volume surgeons and concentrated in higher volume centres then the cost-benefit of UKR may more become more favourable. To produce a safe and sustainable service and improve outcomes for patients, consideration should be given to whether this service should be centralised with fewer centres and surgeons undertaking a higher volume of procedures, mirroring the restructuring of other surgical specialties (NHS specialised services 2011).

This analysis indicates a threshold value for centre and surgeon volume of 100 procedures over the 8 years of observation, equating to approximately 13 procedures per year. However, this figure is likely to be a conservative estimate. A centre/surgeons volume is likely to be greater than that recorded for this analysis owing to the fact that we only analysed medial cemented Oxford implants performed for osteoarthritis and there may have been a failure to report procedures in the early years of the registry. Reported volume thresholds for TKR are higher with suggested minimum volumes of 50-100 TKRs per surgeon per year (Norton 1998, Katz 2004). Previous analysis of UKR and centre volume demonstrated better outcomes for Oxford UKR if >23 were performed per year (Robertsson 2001). As such a threshold of 13/year could be considered a minimum value for the number of medial cemented Oxford UKRs performed annually. Surgeon volume would also appear to be more important than centre volume and caution is advised if centres are undertaking >13 UKR/year based on the combined input of several low volume surgeons. Altogether 303 of the 366 centres (82.8%) and 868 of the 919 surgeons (94.5%) performed less than 100 primary cases, emphasising the large numbers doing inadequate volumes.

Analysis 4: Summary

The UKR debate has previously focussed on if and when these procedures should be performed. Specialist, high volume centres and surgeons produce superior results so a more pertinent question may be to ask “where should these procedures be performed?”. Improving the cost effectiveness of UKR relies on revision rates and functional outcomes

that approach those seen after TKR. One way to achieve this would be to concentrate the UKR workload, allowing centres and surgeons to undertake the necessary volumes required to improve patient outcomes and drive up standards of care.

Analysis 4: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

The biggest problems encountered when undertaking an analysis such as this are determining how to measure volume and subsequently how to categorise surgeons and centres based upon their volume. There is currently no consensus on the best way to measure surgical volume. Possible approaches to determining volume include using the previous years workload, totaling or averaging out the work undertaken over a pre-specified period, or examining each year over a pre-specified period individually (i.e. surgeons can be high volume one year and low volume the next within the analysis). Operative volume is a very simplistic way of assessing surgical performance. Performance is influenced by a number of factors which are difficult to measure and even more difficult to quantify. Aspects of prior surgical training such as the institution in which training took place, the experience of the trainer, the duration of training, the intensity of training and the individual's ability to be trained are all likely to impact on subsequent performance. These factors may have a significant bearing upon surgical outcome but their role is poorly understood.

For this analysis volume was determined by observing the number of procedures performed over an 8 year period. It was felt this was the best way of measuring volume using the available data as it includes some measure of prior experience rather than simply using the most recent year's volume in isolation. It takes into account unexpected departures from

the usual output, which may misrepresent true volume, due to centres and surgeons being busier or quieter than anticipated during the previous year. It also makes some correction for the newly appointed surgeon who is doing high volumes at the end but may not have built up the wealth of experience of more senior colleagues over a prolonged period. We accept that others may have approached this problem differently. Interest in the effects of surgical volume upon outcome within a range of surgical specialties is increasing. To assist in future work it would be useful to review how volume has been measured in the past and whether there are specific methodological advantages to one method over the rest.

We chose to categorise volume into 5 groups roughly based on the quintile distribution of the data. This allowed us to identify 100 procedures in 8 years (approx. 13/year) as a threshold below which poorer revision rates were observed. The size of the dataset would have allowed us to use a greater number of groups but it was felt this would over-complicate the analysis. If a greater number of groups had been used it might have been possible to refine this threshold further and gain a greater insight into the relationship between surgical volume and revision.

The effects of centre volume are difficult to quantify as they are dependent upon the "pooled" results of a variable number of surgeons with a range of experience. This is emphasised in the Kaplan-Meier plot for centre volume in which the two highest and two lowest groups showed a similar pattern of failure but the intermediate volume group showed a unique failure pattern with an initial high rate of failure which improved over time. This pattern of failure may be partly explained by the observation that in contrast to the low and high volume centres, in which procedures were almost exclusively performed by low and high volume surgeons respectively, the intermediate centre volume group was highly heterogeneous for both surgeon numbers and experience. For this reason one must be cautious when interpreting information relating to centre volume from this and other analyses. This limitation, taken in combination with the finding that surgeon volume was the

most important influence upon revision, suggest that future work in this area should focus primarily upon the impact of individual surgeon volume upon outcome.

Chapter 5: Clinical outcomes after total knee replacement

5.1 Current issues surrounding Total Knee Replacement (TKR)

Total Knee Replacement (TKR) has now surpassed hip replacement as the commonest joint replacement procedure in the world (Kurtz 2007). Approximately 80,000 such procedures are performed in England and Wales each year at an estimated annual cost of £500million (DoH 2012, NJR-AR 2012). The demand for knee replacement is predicted to increase six-fold between 2005 and 2030 as we service an increasingly elderly, yet functionally demanding population (Kurtz 2007). This poses something of a problem for health care commissioners with demand likely to outstrip supply. Allied to this are the significant costs associated with these procedures which must be considered given the fact that they aim to improve quality of life rather than save or sustain it. Despite widespread adoption the reported rates of patient satisfaction with TKR are <85% (Robertsson 2000, Noble 2006, Baker 2007, Bourne 2010) and up to 20% of patients fail to demonstrate improvements in health scores post-operatively (Browne 2007). This places TKR in a precarious position, at the mercy of health purchasers who have to balance the needs of their patients, ever tightening budgetary constraints and a lingering concern about a significant subgroup of patients who remain dissatisfied following surgery.

In the context of these developments it is increasingly important for surgeons and centres performing TKR to demonstrate the procedures they are performing are effective. The operations performed should bring about tangible improvements in function and quality of life. This reflects current NHS reform which aims to “put quality at the heart of everything it does” (Darzi 2008). Patient Reported Outcome Measures (PROMs) offer a way to measure these improvements and are being increasingly used to supplement more traditional methods of assessing outcome such as patient safety and implant survival (Darzi 2008, DoH 2009). They are now an integral part of the evaluation of the clinical quality of providers of elective procedures, and are increasingly being used to benchmark provider performance,

assess the efficacy and cost-effectiveness of different approaches to care, empower commissioners contracting services and drive patient choice (Darzi 2008, DoH 2009).

Another recent development has seen a number of NHS trusts introduce guidance for primary care physicians on the type of patient suitable for referral to secondary care for consideration of TKR. These criteria have focused on limiting access to surgery in patients who are either deemed to not have severe enough disease (based on their pre-operative PROMs), or who have what is felt to be undesirable patient characteristics (smokers, obesity) (NHS North Lincolnshire 2011, NHS Warwickshire 2011). TKR is therefore effectively being rationed based on these factors in an attempt to control budgets and contain NHS spending.

This is a major issue. If we consider the example of obesity we discover that in 2009 22 per cent of men and 24 per cent of women in the United Kingdom were classified as obese (Body Mass Index (BMI) ≥30kg/m^2) (NHS Information centre 2012). The upward trend in the proportion of the population classified as obese is expected to increase further over the next 10 years (Zaninotto 2012). Obesity is an independent causative factor in the development of knee osteoarthritis (Lohmander 2009) and the average patient presenting for TKR is now classified as obese (NJR-AR 2012). Growth in the proportion of the population with obesity, combined with an increased demand for knee arthroplasty as we service an increasingly elderly population, will inevitably lead to a rise in the number of obese patients requesting TKR. Introducing referral barriers based on BMI thresholds will therefore deny TKR to an increasing proportion of the population. It is therefore important that we are certain that these barriers are evidence based.

Historically both obese (BMI ≥30 kg/m^2) and morbidly obese (BMI ≥40 kg/m^2) patients have suffered a higher incidence of complications (Dowsey 2008, Samson 2010) and lower rates of implant survival (5-year survival rates = 74%) following TKR (Amin 2006). This view has,

however, been challenged by contemporary reports of equivalent rates of complications (Suleiman 2012) and mid-term survival (Yeung 2011, Dalury 2012) irrespective of the patient's pre-operative BMI. Two recent reviews on the subject concluded that complication rates are higher in the obese and the morbidly obese, and survivorship is inferior (Dowsey 2008, Samson 2010). However, insufficient evidence is currently available neither to definitively determine the effect on functional and quality-of-life outcomes in obese patients undergoing TKA (Samson 2010) nor to determine if there is a threshold above which the risks of surgery outweigh the benefits. This raises questions about the validity of the barriers to surgery and the arbitrary thresholds on which they are based. It also directly conflicts with NICE guidance that "patient-specific factors (including age, gender, smoking, obesity and comorbidities) should not be barriers to referral for joint replacement surgery." (NICE (CG59)2007). Without further supportive evidence the use of these barriers is highly questionable.

The impact of patient factors and pre-operative PROMs upon post-operative function and patient satisfaction are poorly understood and their discriminative capacity in predicting which patients are likely to be satisfied following TKR has yet to be established. Recent examination of the condition specific metric (Oxford Knee Score (OKS)) recorded within the PROMs demonstrated that it had no predictive accuracy in relation to post-operative satisfaction (Judge 2011). The use of these tools for limiting and/or prioritising access to surgery is therefore also unclear.

Given the current vogue for using PROMs to evaluate the quality of clinical care and also to limit access to surgery it is of paramount importance that the orthopaedic community understands how PROMs are influenced by a range of patient and surgeon related factors. By understanding PROMs clinicians and managers can instigate strategies to optimise outcomes and improve standards. While it is well known that a variety of patient factors influence outcome (Fortin 1999, Baker 2007, Bourne 2010), there may be little the surgeon can do to influence them apart from improved patient selection. Failure to appreciate these

effects could lead certain centres to be unfairly penalised by the case-mix of the populations they serve. However surgeons may mitigate this risk by exercising greater autonomy over a number of factors and techniques employed at the time of surgery

The surgeon has an array of potentially modifiable factors available to them at the time of surgery. He has a degree of choice over the brand of knee replacement he uses and thereafter the specifics of the bearing type, meniscal design, and patella resurfacing. He also decides how the operation will be performed, by what surgical approach and whether specialist equipment such as computer navigation will be employed. The impact of a number of these factors upon functional outcome has been investigated using randomised trial designs previously (Wylde 2008, Burnett 2009, Johnston 2009, Choi 2010). However, these publications have all failed to demonstrate any significant benefit deriving from these surgical factors. The relative influence of surgeon determined factors upon the PROMs outcomes that are now so important remains unclear.

The on-going Knee Arthoplasty Trial (KAT) (HTA ref: 95/10/01, finding £1,381,371) commenced in 1998 and is due for completion in later 2013. It aims to clarify the role of different approaches to surgery by asking 1) Should the inner surface of the knee cap be resurfaced? 2) Should the tibial component have a metal back? and 3) Should the knee replacement have a mobile bearing?. When completed this study will have taken 15 years to answer these questions but will not have addressed other pertinent questions relating to the surgery performed such as is there any role for computer navigation?, is there any benefit to minimally invasive surgical techniques?, does it make a difference if you use a cruciate retaining or a posterior stabilised knee design?, how does the brand of knee replacement and where you have your surgery performed influence post-operative functional outcomes and patient satisfaction?. The results of this trial are therefore likely to answer some of the many questions surgeons are asking but leave them scratching their heads about others.

It may be possible to address a number of the issues raised above using registry data. The NJR and PROMs data available was therefore used in an attempt to answer a number of questions about TKR. These included:

1. Which surgical factors influence improvements in PROMs following TKR and what is their influence relative to patient related factors?
2. How do pre- and post-operative factors influence patient satisfaction and the perception of symptom improvement following TKR?
3. What is the nature of the relationship between pre-operative, post-operative and the improvements in PROMs and obesity?
4. Is there evidence to support the idea of an obesity “cut off” as a basis for withholding surgery based on functional outcomes?

These questions are addressed in the subsequent 3 analyses.

1. Analysis of the influence of surgical factors on PROMs.
2. Analysis of the relationship between pre- and post-operative factors and patient satisfaction with TKR.
3. Analysis of the relationship between Body Mass Index and PROMs.

5.2 Analysis 5: Analysis of the influence of surgical factors on PROMs

Analysis 5: Aims

To determine:

1. Which surgical factors influence improvements in PROMs following TKR.
2. What is the relative influence of surgical factors in comparison to a number of patient related factors.

Analysis 5: Methods

This analysis was performed on the NJR-PROMs dataset described in section 3.4.4. From this dataset data was extracted on all patients for whom the primary indication for TKR was osteoarthritis. In total 22,691 NJR-PROMs records were available for analysis. The demographic, surgical and implant variables available for analysis are listed in table 3.8 (section 3.4.4). These included a number of clinically important operative (lead surgeon grade, hospital type, minimally invasive surgery, computer navigation) and implant variables (meniscal type, bearing type, patella resurfacing, implant brand) whose impact of the improvements in PROMs could be assessed. The improvement in the OKS and EQ5D were again used as the primary outcomes for this analysis.

The standard modeling process described in section 3.6.3 was used for this analysis. Initial analysis to establish the relationship between the individual explanatory variables and magnitude of the OKS/EQ5D improvements was undertaken using one-way analysis of variance (ANOVA) and univariable regression. Multiple regression was then employed to account for the significant variations in case-mix (patients' characteristics), pre-operative scores and implant usage which could potentially confound any comparative analysis. A p value of $p<0.001$ was used to indicate statistical significance given the large number of

patients and variables available. This was done to limit the effects of type I error and help produce a parsimonious regression model. Initial analysis revealed there was a significant amount of missing data for Body Mass Index (BMI) (9100 cases (40%)). While the patients with and without BMI data were reasonably matched for patient factors there were significant differences in the surgical factors (hospital type, lead surgeon grade, navigated surgery, meniscus type, implant brand and bearing type (all $p<0.05$)) between these groups. This suggested that patients with BMI data might not be representative of the total population and that the recording of BMI may be related to the type of surgery performed. As such inclusion of BMI might bias any analysis of the influence of surgical factors upon outcome. Models were therefore produced that did not include BMI as an explanatory variable.

For the final models the model estimates with 95% confidence limits are provided to allow comparison of the adjusted effect size between variables. Estimates effectively represent the predicted changes in the OKS and EQ5D for that variable once the effect of all other variables included in the model are considered. For continuous variables (age, BMI, OKS and EQ5D scores) these estimates relate to the expected changes in the response outcome (OKS/EQ5D) for a unit change in the explanatory variable. For categorical variables the estimates are given relative to the base reference category. Model residuals and other checks of model adequacy were satisfactory. There was one significant 2-way interaction between variables included within the final models which is discussed within the results section.

Analysis 5: Results

Univariable analysis

Surgical factors that influenced the magnitude of the OKS and EQ5D improvements included implant brand and hospital type (Both OKS/EQ5D), and bearing type (OKS only) (all $p<0.001$) (table 5.1). For implant brand there was a noticeably greater improvement in both OKS and

EQ5D with the NexGen implant when compared to other groups (p<0.001) (table 5.2). The effect of hospital type was related to differences between NHS hospitals (20,288 cases) and Independent hospitals (1491 cases)/ Independent Sector Treatment Centres (ISTCs) (912 cases), with greater improvements seen in latter two institutions. There was also an association between the magnitude of the OKS improvement and bearing type which indicated an advantage for fixed over mobile bearing implants. No other surgical factors individually influenced the outcomes for either score.

Predictor variable	Dependent variable	
	OKS improvement	EQ5D improvement
Patient factors		
Age (years)	0.007	0.65
Pre-operative OKS	**<0.001**	**<0.001**
Pre-operative EQ5D index	**<0.001**	**<0.001**
Number of comorbidities	**<0.001**	0.97
Time from operation to post-operative PROMs questionnaire collection (days)	0.45	0.36
Gender	**<0.001**	**<0.001**
Side of surgery	0.58	0.59
Pre-operative disability	**<0.001**	**<0.001**
Pre-operative general health	**<0.001**	**<0.001**
Depression	**<0.001**	**<0.001**
Anxiety level	0.52	**<0.001**
ASA grade	0.007	0.85
Duration of symptoms	**<0.001**	0.007
Surgical factors		
Lead surgeon grade	0.20	0.24
Minimally invasive surgery	0.29	0.84
Computer Navigated	0.08	0.62
Bearing	**<0.001**	0.23
Meniscus	0.51	0.03
Patella Resurfaced	0.77	0.64
Brand type	**<0.001**	**<0.001**
Hospital type	**<0.001**	**<0.001**

Table 5.1: Summary of the p values for univariable analysis with improvement in OKS and EQ5D index as the response variables.

In addition a number of patient factors were seen to influence the PROMs improvements. Most notably there was an inverse relationship between the magnitude of the improvements observed and the corresponding pre-operative PROMs for both the OKS and EQ5D. Patients who start off worse had a greater chance of improving presumably because they have greater scope for improvement. Better overall health (pre-operative general health, ASA code, pre-operative disability, anxiety/depression) was also related to a greater PROMs improvement. These findings were consistent with the associations seen in analysis 1 (section 4.2).

		Pre op	**Post op**	**Change**
NexGen® (n=3283)	OKS	18.9 (95% CI 18.6 to 19.1)	35.1 (95% CI 34.8 to 35.5)	16.2 (95% CI 15.9 to 16.6)
	EQ5D	0.409 (95% CI 0.398 to 0.420)	0.730 (95% CI 0.722 to 0.739)	0.323 (95% CI 0.310 to 0.333)
PFC® (n=8287)	OKS	18.7 (95% CI 18.5 to 18.8)	33.9 (95% CI 33.7 to 34.1)	15.2 (95% CI 15.0 to 15.4)
	EQ5D	0.396 (95% CI 0.389 to 0.403)	0.708 (95% CI 0.702 to 0.714)	0.312 (95% CI 0.305 to 0.320)
Genesis 2® (n=1818)	OKS	19.3 (95% CI 18.9 to 19.6)	33.6 (95% CI 33.2 to 34.1)	14.3 (95% CI 13.9 to 14.8)
	EQ5D	0.425 (95% CI 0.410 to 0.440)	0.707 (95% CI 0.694 to 0.719)	0.282 (95% CI 0.266 to 0.299)
AGC® (n=2398)	OKS	19.4 (95% CI 19.1 to 19.7)	34.2 (95% CI 33.8 to 34.6)	14.9 (95% CI 14.5 to 15.2)
	EQ5D	0.429 (95% CI 0.416 to 0.442)	0.715 (95% CI 0.704 to 0.726)	0.286 (95% CI 0.272 to 0.300)
Triathlon® (n=1896)	OKS	19.6 (95% CI 19.2 to 19.9)	34.4 (95% CI 34.0 to 34.9)	14.8 (95% CI 14.4 to 15.3)
	EQ5D	0.436 (95% CI 0.422 to 0.451)	0.725 (95% CI 0.713 to 0.737)	0.289 (95% CI 0.274 to 0.304)
Other (n=5009)	OKS	19.1 (95% CI 18.9 to 19.3)	33.6 (95% CI 33.3 to 33.9)	14.5 (95% CI 14.2 to 14.8)
	EQ5D	0.407 (95% CI 0.398 to 0.416)	0.702 (95% CI 0.694 to 0.709)	0.294 (95% CI 0.285 to 0.304)
Total (n=22,691)	OKS	19.0 (95% CI 18.9 to 19.1)	34.1 (95% CI 33.9 to 34.2)	15.1 (95% CI 14.9 to 15.2)
	EQ5D	0.409 (95% CI 0.405 to 0.414)	0.712 (95% CI 0.708 to 0.716)	0.303 (95% CI 0.298 to 0.307)

Table 5.2: Improvement in the OKS and EQ5D scores analysed by implant brand.

Multiple regression analysis

Once the influence of other variables was considered the only surgical factors seen to influence outcome were implant brand and hospital type (table 5.3). The effect of bearing type (p=0.003) did not appear within the final OKS model as it did not reach the criteria (p<0.001) for inclusion. The NexGen implant remained the best performing implant within these models. The variation in PROMs ascribed to brand type is given in the estimate columns of table 5.3. The difference between the best (NexGen) and worst ("Others") brands within these models was 1.7 (95%CI 1.4 to 2.2) for the OKS and 0.034 (95%CI 0.023 to 0.045) for the EQ5D. The statistical significance seen between brands has been facilitated by the large numbers available for analysis and the associated precision of the estimates for the mean PROMs improvement. However, the variation in improvement for each brand remains large with significant overlap in the distribution of improvements between the different brands (figure 5.1).

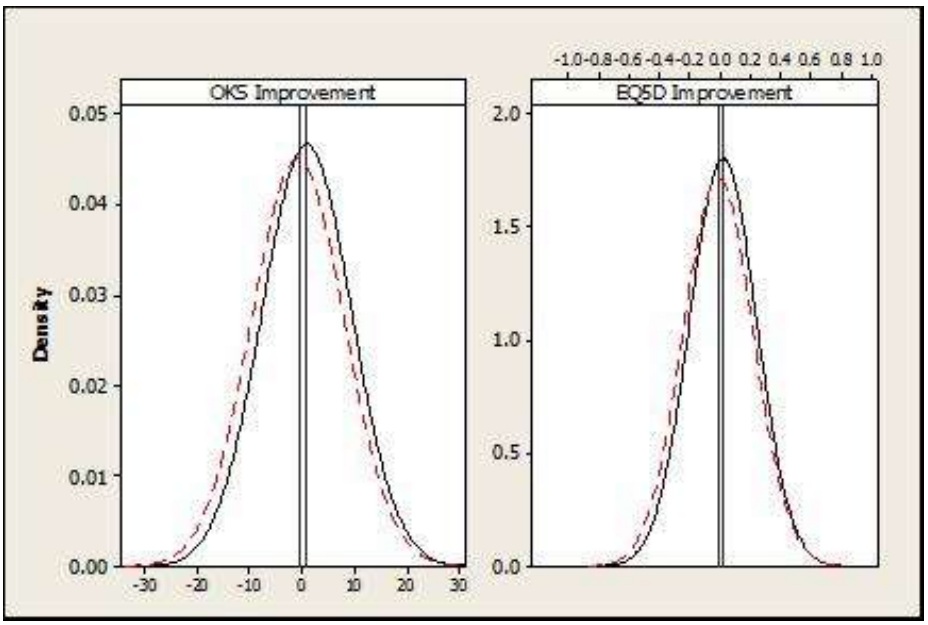


Figure 5.1: Distributions of the improvement in OKS and EQ5D after adjustment using the regression models for the best (NexGen®: Black) and worst (Others: Red) brands. The reference lines represent the mean for each brand in relation to the overall mean for all patients (zero line on horizontal axis).

For hospital type the improvements in both OKS and EQ5D were again greater for ISTCs when compared NHS hospitals (OKS 1.8 (95% CI 1.2 to 2.5), EQ5D 0.040 (95%CI 0.024 to 0.057)). There was no significant difference between independent hospitals and ISTCs. The only significant interaction between either surgical factor with other variables in the final models was between depression and implant type for the magnitude of the EQ5D improvement. However, on further graphical investigation the effect of this interaction was not found to be important.

While both brand and hospital type were seen to influence PROMs improvements the effect sizes of these factors were modest compared to those attributable to a number of patient factors. The effect sizes (the predicted effect of each variable on the magnitude of the change in PROMs) for all of the variables included in the final models are given as the model estimates for that variable in table 5.3. Within both models the most important variables were the relevant pre-operative PROMs score, the patient's ratings of their pre-operative general health, and the presence of anxiety and depression. For example, table 5.3 shows that a for every one point increase in the pre-operative OKS the change in OKS is predicted to decrease by -0.66 points, thus an increase in the pre-operative OKS of 10 points predicts a reduction of -6.6 points in the OKS improvement. The corresponding size effects for patients with the best and worst general health is 8.1 points. This puts into context the differences seen between the best and worst implant brands (1.7) and hospital types (1.8). It must, however, be noted that the effect of each variable on the predicted improvement in PROMs are additive, meaning the effects of brand and hospital type are the same irrespective of changes in the other variables.

	OKS improvement			EQ5D improvement		
Predictor Variable	**Estimate**	**95% CI**	**p value**	**Estimate**	**95% CI**	**p value**
Patient variables						
Age	0.06	0.04 to 0.07	<0.001	0.002	0.002 to 0.002	<0.001
Pre-operative OKS	-0.66	-0.67 to -0.64	<0.001	0.003	0.003 to 0.003	<0.001
Pre-operative EQ5D index	-	-	-	-0.891	-0.905 to -0.876	<0.001
Number of Co-morbidities	-0.25	-0.39 to -0.12	<0.001	-	-	-
Pre-operative disability						
No	Reference			Reference		
Yes	-1.49	-1.75 to -1.23	<0.001	-0.049	-0.056 to -0.042	<0.001
Pre-operative general health						
Excellent	Reference			Reference		
Very Good	-1.12	-1.78 to -0.45	<0.001	-0.038	-0.056 to -0.020	<0.001
Good	-2.78	-3.42 to -2.12	<0.001	-0.076	-0.093 to -0.058	<0.001
Fair	-5.23	-5.93 to -4.53	<0.001	-0.152	-0.171 to -0.133	<0.001
Poor	-8.13	-9.09 to -7.16	<0.001	-0.243	-0.269 to -0.217	<0.001
Depression						
No	Reference			Reference		
Yes	-0.95	-1.44 to -0.46	<0.001	-0.066	-0.079 to -0.054	<0.001
Anxiety level						
No anxiety/depression	Reference			Reference		
Moderate anxiety/depression	-1.17	-1.45 to -0.90	<0.001	-0.041	-0.048 to -0.033	<0.001
Severe anxiety/depression	-2.78	-3.48 to -2.07	<0.001	-0.097	-0.117 to -0.077	<0.001
ASA grade						
Grade 1	Reference			Reference		
Grade 2	-0.27	-0.67 to 0.13	0.19	-0.011	-0.021 to -0.000	0.046
Grade 3/4	-1.00	-1.52 to -0.49	<0.001	-0.037	-0.051 to -0.024	<0.001
Surgical variables						
Brand type						
NexGen®	Reference			Reference		
PFC®	-0.98	-1.35 to -0.62	<0.001	-0.015	-0.025 to -0.005	0.003
Genesis 2®	-1.50	-2.02 to -0.98	<0.001	-0.023	-0.037 to -0.009	0.001
AGC®	-1.20	-1.68 to -0.72	<0.001	-0.023	-0.036 to -0.010	<0.001
Triathlon®	-0.94	-1.46 to -0.43	<0.001	-0.011	-0.024 to 0.003	0.11
Other	-1.74	-2.16 to -1.36	<0.001	-0.034	-0.045 to -0.023	<0.001
Hospital Type						
NHS Hospital	Reference			Reference		
Independent Hospital	0.83	0.35 to 1.31	0.73	0.017	0.004 to 0.029	0.01
ISTC	1.84	1.23 to 2.45	<0.001	0.040	0.024 to 0.057	<0.001

Table 5.3: Summary of the variables significant in the final multiple linear regression models. Estimators with 95% CI are presented to allow comparison of the effect size of each variable. For categorical variables the estimators are given as differences relative to the first category. For continuous variables the estimate describes the effect on the change in score if the predictor variable was increased by one point.

Analysis 5: Discussion

This analysis found that of the variables considered the only surgical factors influencing PROMs improvements were implant brand and hospital type. However, the effects attributable to these factors were small in comparison to those associated with a range of patient factors whose influence upon PROMs improvements was more pronounced.

PROMs are becoming an established means of assessing and comparing Trusts and surgeons. In the current climate of transparency and accountability, results of the NJR and PROMS are within the public domain. Patients are becoming far more educated, particularly with the ability to access internet resources and are thus becoming more selective with regard to who will perform their joint replacement (Marcario 2003). This environment provides the motivation to drive quality improvement, track productivity and foster inter-unit competition. In addition PROMs data is soon to be used for ranking, re-imbursement and the identification of outliers with individual surgeon's outcomes displayed alongside other surgeons' performing the same procedure (Fairley 2008). Consequently, these measures wield potentially substantial financial weight. Should PROMs data be used for benchmarking and comparison, any advantage (even theoretical) becomes desirable, regardless of change in patient function. However, incentivising institutions and individuals based on outcome measures becomes more difficult to defend if the gain in clinical benefit is unclear (Kay 2011).

A number of previous studies have assessed the impact of surgical factors upon functional outcomes. Surgical factors studied include patella resurfacing (Nizard 2005, Burnett 2009), bearing type (mobile vs. fixed bearing) (Jacobs 2004, Wylde 2008), tibial component design (metal backed vs. all polyethylene) (Najibi 2003), meniscus type (cruciate retaining vs. posterior stabilised) (Tanzer 2002), computer navigation (Bauwens 2007, Luring 2011), minimally invasive surgery, high flexion knee systems (Choi 2010), gender specific knee systems (Kim 2010) and implant type (Koskinen 2010, Scott 2010, Endres 2011). None of

these has demonstrated a clear functional advantage. Implants and their corresponding design traits are surrounded by a massive advertising framework and a strong representative presence. It can be hard to separate out performance in terms of patient outcome measures due to the number of confounding variables that accompany the varied patient mix. The NJR provides a wealth of information with which we have been able to examine the impact of surgical factors, after adjustment for important patient factors, on PROMs improvements.

Papers specifically comparing implant brands have failed to show improved outcomes for one particular brand (Scott 2010, Endres 2011). However, these analysis are based on small numbers and limited data. This analysis has the benefit of large numbers (n=22,691) and a robust statistical analysis, allowing for the identification of significant relationships unappreciated in smaller analyses.

Previous reports have suggested that ISTC's achieve higher patient satisfaction rates when compared to NHS institutions (97.5% *versus* 89%) (DoH 2007). A thorough evaluation of their performance, quality of care and equality of access has not been possible due to poor compliance with data submission to the Hospital Episode Statistics (HES) database and a lack of enforcement of this contractual requirement. The Healthcare Commission of 2007 did, however, note ISTCs had fewer post-operative readmissions and shorter lengths of stay (DoH 2007), reflecting the case-mix for ISTCs which is self-selecting and by its very nature, excludes patients with complex co-morbidities. This analysis made adjustments for case-mix and as such the observed differences in PROMs cannot be attributed to case-mix variation. The findings may, however, be partly due to the inclusion of one high volume ISTC with particularly good results which contributed 44% (402 of 912 cases) of the procedures for the ISTC group (11 ISTCs). This is likely to bias the results for ISTCs, with the results reflecting outcomes for this institution rather than for ISTCs as a whole.

While surgical factors were important in the final models, the factors demonstrating the greatest influence on PROMs outcomes were predominately patient related. This is consistent with a number of other studies that have demonstrated significant associations between outcome and a variety of patient factors (Fortin 1999, Baker 2007, Bourne 2010). In all models the most important and consistent predictor of outcome was the relevant pre-operative PROMS score. Better outcomes were also seen in patients with the best levels of pre-operative general health, disability and depression/anxiety, which is consistent with the literature (Lingard 2004, Chang 2010). While the best post-operative scores were seen in patients with the best pre-operative scores, patients with lower pre-operative scores demonstrated the greatest improvements. This observation is likely related to the larger potential improvements available to these patients and PROMs ceiling effects due to the inability for these scores to adequately detect top end differences. The absolute post-operative score is thus arguably an unfair measure, as patients with low pre-operative scores have lower post-operative scores and as such are unlikely to achieve pre-determined target scores designated as the bookmark of success. Instead, the overall change in score, adjusted for the pre-operative starting point, should be used a more accurate reflection of the patient's experience.

Difficulty lies in establishing the relationship between the amount of change required in the scoring system and the perceptible clinical difference. For a difference to be important, it has to make a difference to the patient. In many units, it has become part of the usual surgical pre-assessment and follow-up to complete standardised and validated scores. The minimal clinically important difference (MCID) for the OKS has not been published, but is has been demonstrated that the MCID for TKR using the Western Ontario and McMaster Universities Osteoarthritis index (WOMAC) score is 15 points and the Short Form 36 (SF-36) at least 10 points (Escobar 2007). It is unlikely therefore that the differences of less than 2 OKS points seen between implants and institutions will be clinically relevant. This demonstrates that patient factors are more important than anything we as surgeons can control at the time of surgery. It is, however, differences of this magnitude that are

monitored by managers and politicians and considered as evidence of performance with far-reaching implications for both the surgeon and their organisation.

These findings highlight the fundamental importance of adequate case-mix adjustment when comparing institutions and surgeons. Without adequate correction surgeons could potentially become PROMs outliers dependent solely upon the population they serve and the cases they undertake. In a climate of financial incentives, failure to recognise and make allowance for these factors could result in surgeons cherry-picking cases to obtain the best results. The danger could be that patients with significant co-morbidities, poor mental health scores or reduced functional ability for whom arthroplasty could potentially be life-changing, may be denied surgery. Therefore, it is vital that discussion of PROMs for payment by results includes detailed analysis of case-mix and pre-operative scores before the less tangible effect of the individual surgeon's skills can be teased out. Whilst the surgeon can influence the implant choice and timing of surgery, he/she has little sway on case-mix and like it or not, the outcome, or change in outcome is what the commissioning bodies will use to determine workload choice.

Analysis 5: Summary

This analysis found that of the surgical factors analysed only implant brand and hospital type were associated with PROMs improvements. However the effects of these categorically collected surgical factors were small when compared to patient factors, and in particular the pre-operative PROMs and general health status. The influence attributable to the pre-operative PROMs also highlights the importance of collecting both pre and post-operative data when comparing groups, and the significance of the change in score *versus* the absolute outcome score. Given these findings, it would seem to be inappropriate to base decisions about the effectiveness of surgical interventions on post-operative scores alone.

Analysis 5: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

The 'surgical factors' used for this analysis were those that could be extracted from the NJR dataset. Implant coding within the NJR dataset is complex. In many cases there are greater than 20 columns of data relating to each implant which makes it difficult to effectively assimilate implant data. We therefore had to simplify the way in which we grouped surgical data particularly for factors such as implant brand, meniscus and bearing. Given more time and a greater understanding of the coding structure it might have been possible to extract a greater level of detail for these factors.

This analysis suffered the same problem as analysis 1, namely that the final models produced only explained a small amount of the observed variation in the response variables (OKS and EQ5D improvement). The influence and significance of variables within the regression models naturally change as variables are added or removed. The addition of information relating to 'missing' variables such as patient expectation or mental health might change the model so that variables that were initially significant are no longer so. This is concerning, particularly when there is a commercial interest in the results of the analysis due to the inclusion of implant brand data.

The minimally clinically important difference (MCID) for the OKS is not yet known. It is, however, likely to be greater than the size effects of many of the variables included in the final regression models. A clinical study examining, for example, the influence of implant brand upon the OKS would use information on the distribution of this measure in combination with the MCID to perform a power calculation and sample size analysis. Recruitment would be based on this analysis and would not exceed it. Due to the size of the

NJR-PROMs dataset the analysis has been able to identify variables that have an influence at a level below the MCID. Reporting these associations relies on the readership understanding the concept of the MCID and applying the findings within this context. Unfortunately this is not always the case as many clinicians focus primarily on high significance (p values) even if it is associated with a marginal, clinically irrelevant outcome. Given these concerns it may have been better to model and report this data based not only on the p value but also on the model estimate, with variables only included if they reached statistical significance and their model estimate was in excess of the MCID. If these criteria were applied to this analysis only the pre-operative OKS, pre-operative general health and possibly the anxiety level would have made it into the final model for the OKS improvement. For the EQ5D improvement only the pre-operative EQ5D score would have made it into the final model. These simplistic models highlight only the most important explanatory variables and help avoid over-interpretation.

5.3 Analysis 6: Analysis of the relationship between pre- and post-operative factors and patient satisfaction with TKR

Analysis 6: Aim

1. To establish how pre- and post-operative factors influence patient satisfaction and the perception of symptom improvement following TKR.

Analysis 6: Methods

This analysis was performed on the NJR-PROMs dataset described in section 3.4.4. From this dataset data was extracted on all patients for whom the primary indication for TKR was osteoarthritis. In total 22,691 NJR-PROMs records were available for analysis. The primary outcome for this analysis was the patient's rating of their surgery ("patient satisfaction") described in section 3.5.4. Within this scale the first category was the best response and the fifth category the worst response meaning the scale was inversely related to the OKS and EQ5D scores. A satisfaction rating was available for 22,373 (99%) of the 22,691 patients analysed.

A structured literature review was undertaken to determine which factors had previously been shown to influence satisfaction following knee replacement. All of the identified reports focused solely on the direct relationships between a variety of explanatory factors and satisfaction without consideration for any interactions between these factors. Reported associations with satisfaction included both demographic characteristics / other preoperative conditions (patient age (Noble 2006, Baker 2007), gender (Ethgen 2004, Baker 2007), underlying diagnosis (Bullens 2001, Baker 2007), a lower level of education (Bourne 2010), increasing Body Mass Index (BMI) (Bourne 2010), mental health status/depression (Anderson 1996, Scott 2010, Blackburn 2012), general health status (Anderson 1996, Noble 2006)) and factors measured postoperatively (need for revision surgery (Hawker 1998,

Robertsson 2000, Bourne 2010), post-operative patient reported outcome scores/symptom improvement (Hawker 1998, Baker 2007, Bourne 2010) and fulfillment of patient expectations (Bourne 2010)). Descriptions of the factors available for this analysis are given in table 5.4.

Predictors	***Details***
Age (Years)	
Gender	- Female / Male
Diagnosis	Analysis restricted to only those procedures performed for osteoarthritis as this group comprises >98% of all knee replacements performed. Other indications excluded
Need for revision surgery	PROMs recorded at 6 months post-operatively. The number of revisions at this time-point is small limiting the ability to analyse this variable
Level of education	Not recorded in the NJR-PROMs dataset
Body Mass Index (BMI)	Data only available for 9,874 (42%) patients. Patients with missing BMI data differed from those with recorded BMI data in relation to reported levels of satisfaction and success. This data was therefore excluded from the analysis.
Mental health status/ Depression	Patient reported. Indicates whether the patient has previously been given a diagnosis of depression - No / Yes
Pre-operative General Health	Patient reported: Indicates the patients perception of their own general health with 5 ordered options: - Excellent - Very good - Good - Fair - Poor
Patient reported outcomes **Euroqol index (EQ5D)** **Oxford Knee Score (OKS)** **Operative Success** The patients perception of surgical success	 Described in section 3.5.3 (Pre-operative, post-operative and change in score) Described in section 3.5.3 (Pre-operative, post-operative and change in score) Described in section 3.5.4 “Overall, how are the problems now in the knee on which you had surgery, compared to before your operation?” - Much better - A little better - About the same - A little worse - Much worse
Patient expectation	Not recorded in the NJR-PROMs dataset

Table 5.4: Explanatory variables identified from the structured literature review and their availability within the NJR-PROMs dataset.

Initial tabular and graphical summaries of the relationship between satisfaction and each of the explanatory variables were supplemented by ordinal logistic regression and structural equation modelling. Ordinal logistic regression was used to analyse the effects of the explanatory variables individually and in combination as satisfaction was measured using a 5 point ordinal scale. Structural equation modelling was then used to investigate the direct and indirect effects of the different factors upon patient satisfaction. Logistic regression models were constructed to assess the relationship between variables available both pre-operatively (age, gender, depression, general health status, pre-operative OKS/EQ5D) and post-operatively (post-operative OKS/EQ5D, 'operative success') with satisfaction. Model identifiability was assessed on the basis of the condition number of Hessain of $<10^4$. Where the condition number of Hessain was in excess of this value the models were rejected on the basis of being ill-conditioned. Initial models including both the OKS and EQ5D in addition to the other patient characteristics were rejected as they did not converge. Models with the OKS alone were also inadequate for the same reason. A resampling exercise using 200 random subsets of 94% of the data showed that all models including the OKS were not identifiable as the condition number for all replicates was in excess on 10^5 (mean 6.5×10^5 SD 4332). However as the OKS and EQ5D were correlated (Preoperative OKS:EQ5D Spearman rank r=0.70, Post-operative OKS:EQ5D Spearman rank r=0.76), and prior analysis had demonstrated the pre-operative OKS and satisfaction were poorly correlated (Spearman Rank correlation 0.04 (95%CI -0.01 to 0.08)) (Judge 2011) a decision was made to proceed using only the EQ5D in addition to the other patient characteristics. Variables were removed using a two-stage stepwise reduction until only significant variables were left in the model.

As the explanatory variables have the potential to influence each other structured equational modelling (SEM) was used to investigate their direct and indirect effects on the satisfaction outcome. SEM is described in section 3.6.4. In summary it is a method for investigating the impacts of variables in systems where there are multiple pathways to the final effect. The approach is used to challenge a hypothetical representation of the system pathways using data. In the present context we hypothesised that patient satisfaction is

driven by patient characteristics, pre- and post-operative function and the perception of symptom improvement ('operative success'). For example, this model is able to recognise that post-operative function is dependent upon preoperative function which in turn is dependent on patient age and gender. Thus there are a set of driving variables that have both direct and indirect effects on satisfaction that are themselves related. The pathway of hypothesised effects for this analysis is shown in figure 5.2. Males and females patients were modeled separately as a grouping index in the analysis.

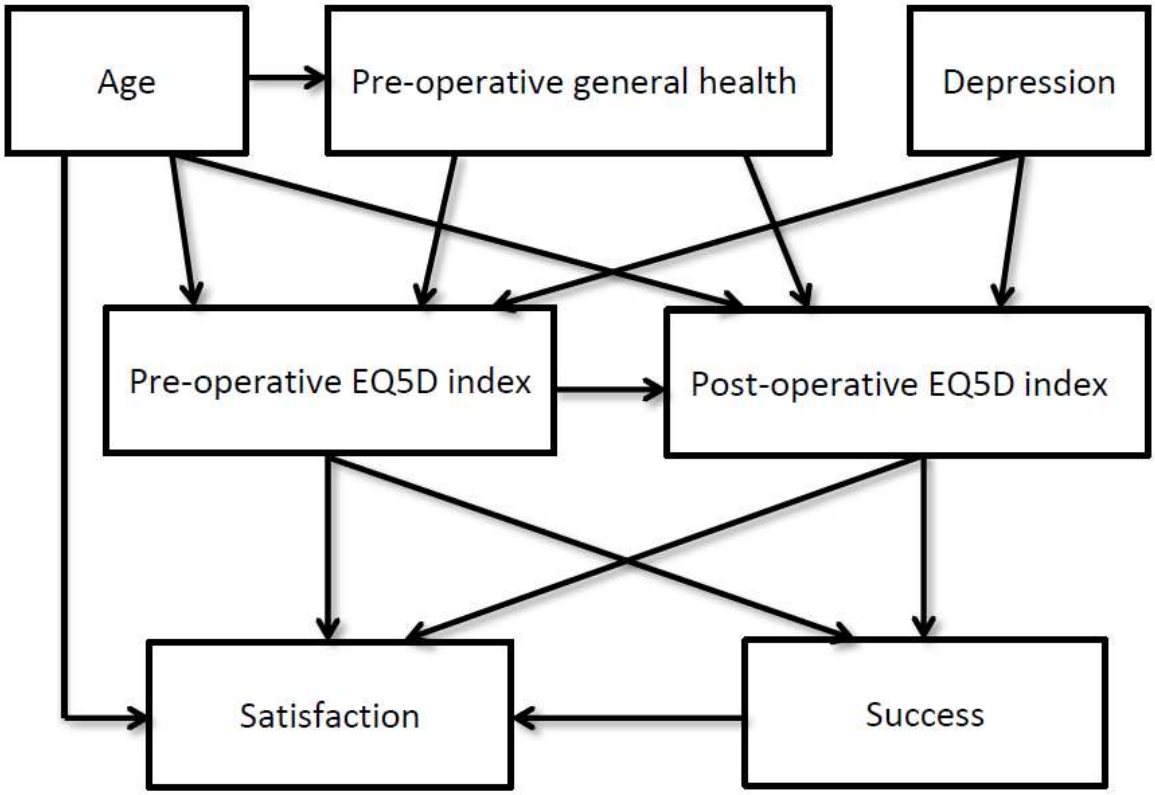


Figure 5.2: Hypothetical model developed to explain the relationship between the available explanatory factors and satisfaction and success

Analysis 6: Results

Initial examination of the relationship between satisfaction and 'operative success' demonstrated that 15,882 (71%) patients perceived their knee symptoms to be 'much better' following surgery. However despite symptomatic improvement only 4,959 (22%) gave their surgery the top satisfaction rating of 'excellent'. Instead the majority of patients rated their surgery as either 'very good' or 'good' (table 5.5). In total there were 1,592 (7%) patients who rated their surgery as 'fair' or 'poor' despite reporting symptomatic improvements (red shaded). This group were on average a year younger (mean age: 68.8 vs. 69.7 years), had higher levels of depression (11% vs. 7%) and 'Fair/Poor' general health (43% vs. 24%), and had lower pre-operative knee (mean OKS score: 16.9 vs. 19.0) and general health (mean EQ5D score: 0.329 vs. 0.407) scores than the total population undergoing TKR. There was also a smaller group of 40 (0.2%) patients who reported high levels of satisfaction despite a worsening of their symptoms (blue shaded).

Satisfaction	Success					
	Much better	A Little better	Much the same	A little worse	Much worse	**TOTAL**
Excellent	4,886	59	3	7	4	**4,959 (22%)**
Very Good	7,344	474	36	18	11	**7,883 (35%)**
Good	3,374	2,002	313	119	23	**5,831 (26%)**
Fair	272	1,251	646	476	122	**2,767 (12%)**
Poor	6	63	111	199	459	**838 (4%)**
TOTAL	**15,882 (71%)**	**3,849 (17%)**	**1,109 (5%)**	**819 (4%)**	**619 (3%)**	**22,278 (100%)**

Table 5.5: Distribution of satisfaction and success following total knee replacement.

Ordinal logistic regression

Patient satisfaction was significantly and positively related to 'operative success'. As the success in operative outcome declined, the odds ratio of a poor satisfaction score increased. A 'much worse' operative outcome had an odds ratio of 845 (95%CI 652 to 1097) of a low satisfaction score relative to a case with 'much better' operative success. Similarly, poor health had a higher odds ratio of low satisfaction (3.3, 95%CI 2.6 to 4.2) relative to a patient in good health. Satisfaction was also related to the pre- and post-operative function as represented by the EQ5D scores (OR = 1.52, 95%CI 1.38 to 1.68 and OR = 0.09, 95%CI 0.08 to 0.10 respectively). In addition being male led to a lower risk of dissatisfaction relative to being female (OR = 0.86. 95%CI 0.81 to 0.91), as did not having a previous diagnosis of depression (OR = 0.83, 95%CI 0.74 to 0.92) (table 5.6). Age was not a significant predictor of satisfaction within the ordinal regression model. The regression diagnostics indicated that post-operative variables had a larger influence upon satisfaction than preoperative variables. For the regression models proportionality assumptions were broadly met except where the number of cases in a class was low, specifically the 'much worse' success group.

Structured equational modelling (SEM)

Full models including all hypothesised explanatory variables and pathways did not adequately describe the variation in satisfaction observed in the data. Z values for all variables included in the final SEM path model (figure 5.3) were significant at $p<0.001$ with the exception of the pathway from age to satisfaction in females where significance was marginal ($Z=1.81$, $p=0.07$). The variable was kept in the final model however, as it was more significant for males ($Z=2.60$, $p=0.009$). Whilst these data show levels of significance for key covariates the coefficients are not standardised in relation to each other so are difficult to compare across pathways. Standardised coefficients for the best fit models for males are shown in figure 5.3. The coefficients on the path lines represent the proportional change in standard deviation of the response in relation to a unit standard deviation change in the hypothesised driving variable.

Covariate	Estimate	SE of Estimate	Z	p	Odds Ratio	Lower 95%CI	Upper 95%CI
Pre-operative variables							
Gender (Reference: Female)							
Male	-0.15	0.028	-5.4	<0.001	0.9	0.8	0.9
Pre-operative EQ5D	0.42	0.049	8.5	<0.001	1.5	1.4	1.7
Depression (Reference: Diagnosis of depression)							
No diagnosis of depression	-0.19	0.056	-3.4	<0.001	0.8	0.7	0.9
General health (Reference: Excellent)							
Very Good	0.67	0.088	8.5	<0.001	2.0	1.6	2.3
Good	1.07	0.078	13.7	<0.001	2.9	2.5	3.4
Fair	1.28	0.082	15.5	<0.001	3.6	3.1	4.2
Poor	1.20	0.116	10.4	<0.001	3.3	2.6	4.2
Post-operative variables							
Success (Reference: Much better)							
A little better	2.79	0.048	58.6	<0.001	16.3	14.8	17.9
About the same	4.10	0.078	52.3	<0.001	60.3	51.8	70.3
A little worse	4.98	0.096	51.7	<0.001	145.5	120.5	175.6
Much worse	6.74	0.133	50.7	<0.001	845.6	651.5	1097.4
Post-operative EQ5D	-2.39	0.075	-31.9	<0.001	0.1	0.1	0.1

Table 5.6: Regression diagnostics for the ordinal regression analysis for the satisfaction response. Note: Caution is advised when interpreting odds ratios >10 such as those seen within the success group.

The final models and model path coefficients for females were similar to those for the males and are therefore not shown. The Root Means Square Error of Association (RMSEA) for the TKR model was 0.064; the Comparative Fit Index (CFI) was 0.97 and the Bollen-Stine bootstrap probability (PBS) for this model was 0.0 indicating that the model was a very good fit for the data. Patient satisfaction was strongly dependent on 'operative success'. Higher levels of satisfaction and 'operative success' were related to higher levels of post-operative function as represented by the EQ5D score. Pre-operative function had a positive impact on postoperative function and both pre-operative function and age had slight but significant impacts on satisfaction, with older patients being more satisfied. In the context of these relationships the level of pre-operative general health and a diagnosis of depression were not found to be important.

The direct and indirect contributions of pre-operative and post-operative covariates on patient satisfaction were estimated by the product coefficient method (table 5.7). Thus, the indirect effect of the post-operative EQ5D on satisfaction was the sum of the products of coefficients for its relationship with 'success' and that for 'success' with satisfaction (i.e. 0.55×0.57) (figure 5.3). Using these method pre-operative variables (age and pre-operative EQ5D) contributed little direct or indirect effect up on patient satisfaction. In contrast, the post-operatively measured 'success' and EQ5D score contributed much more to the perceived satisfaction, with a unit change in 'success' resulting in a >0.55 unit change in patient satisfaction for both males and females. These results further demonstrate the greater influence of post-operative variables upon satisfaction when compared to preoperative variables and highlight the significant role indirect effects have upon this outcome.

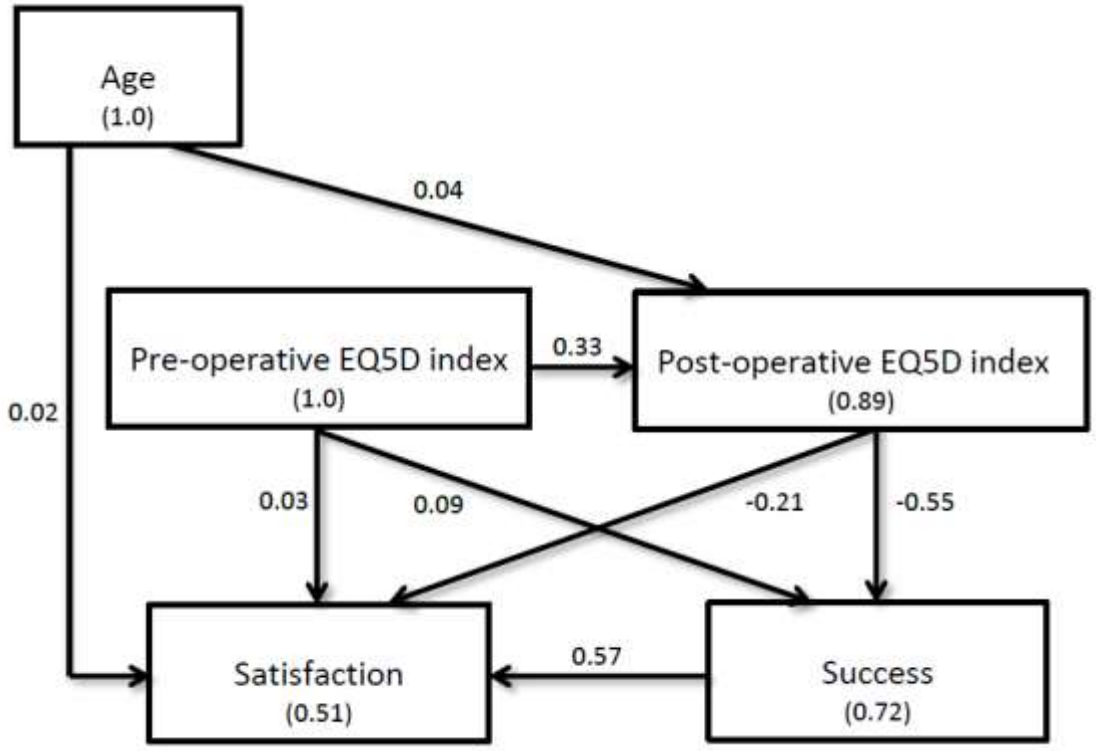


Figure 5.3: Final SEM model for male total knee replacement patients after removal of all non-significant variables and pathways. Values in () represent the proportion of the observed variability which is explained by the suggested model.

	Male		Female	
	Direct	Indirect	Direct	Indirect
Pre-operative variables				
- Pre-operative EQ5D	0.029	-0.115	0.033	-0.126
- Age	0.018	-0.024	0.014	-0.061
Post-operative variables				
- Post-operative EQ5D	-0.214	-0.315	-0.259	-0.313
- Success	0.573		0.552	

Table 5.7: Direct and indirect pathways to effect from the SEM analysis. The values represent the impact of units standard deviation change in the driving variables on the patient's satisfaction with total knee replacement.

Analysis 6: Discussion

This analysis examined the relationships, interactions and predictive capacity of pre- and post-operative variables on post-operative patient satisfaction. While preoperative variables such as gender, a diagnosis of depression, pre-operative general health status and pre-operative EQ5D scores were related to patient satisfaction they were not as strongly related as a number of post-operative factors such as the post-operative EQ5D score and the perception of operative success when both their direct and indirect effects upon satisfaction were modelled.

By using the NJR- PROMS data, with a cohort of patients in excess of 22,000, we have been able to use analytical approaches that are not suitable for the analysis of smaller datasets. This analysis has quantified the relationships between satisfaction, operative success, functional capacity and patient characteristics. It found that only 22% of patients gave the top 'excellent' rating when asked to describe the results of their knee replacement. This occurred despite 71% of patients perceiving their knee to be 'much better' following surgery. This demonstrates the inherent differences between satisfaction and operative success. Success measures the patient's perception of whether they have symptomatically improved following surgery whereas satisfaction measures the extent to which they are

happy with this improvement. Patient expectation is a key determinant of satisfaction (Bullens 2001, Bourne 2010). There were a significant number of patients (1,592) who were dissatisfied despite reporting symptomatic improvements, possibly because their expectations had not been fulfilled. This group was on average younger, had poorer mental and physical health and had poorer pre-operative function that the total population undergoing knee replacement, a finding that is consistent with previous studies analysing the direct effects of these factors upon satisfaction (Anderson 1996, Hawker 1998, Noble 2006, Baker 2007, Bourne 2010, Scott 2010, Blackburn 2012). It is likely that within this analysis 'operative success' was a surrogate marker for the fulfillment of expectation. This may explain why this variable was such a strong predictor of satisfaction and confirms the important role patient expectation has upon outcome.

The modeling techniques employed, allowed us to examine the combined direct and indirect effects of factors upon satisfaction. These approaches confirmed that primary determinant of a patient's level of satisfaction was their post-operative perception of whether their operation was a success. While a number of other factors were also significant in the final models their effects upon satisfaction occurred through an indirect influence upon the 'operative success' variable. In effect this means that these 'preliminary' factors have a greater role in determining whether a patient perceives their surgery as successful, which then in turn determines their level of satisfaction, rather than any effect directly upon satisfaction itself. The findings of the ordinal regression and SEM differed in respect to the presence of age, depression and general health which were significant in one model but not the other. This probably reflects the interdependence between variables. While gender was significant in the ordinal regression, with females having lower levels of satisfaction than males in keeping with previous analyses (Ethgen 2004, Baker 2007), the SEM analysis which was performed separately for males and females demonstrated that the pathways and variables influencing satisfaction are similar irrespective of gender.

The pre-operative OKS has previously been shown not to predict post-operative satisfaction (Judge 2011). Similarly, the analyses based on pre-operative variables alone did not adequately explain the variability in the satisfaction outcome and as such would not be suited to making predictions. It was only when both pre- and post-operative variables were considered together that any prediction of satisfaction could be made. There is clearly merit in attempting to predict which patients are likely to exhibit high levels of satisfaction following surgery to help support patient selection and clinical decision making. However, the findings in this study suggest that this process is complex and simplistic interpretations of small datasets should be treated with caution. Any predictions of satisfaction based on pre-operative variables alone are likely to be heavily modified by the outcome of surgery. This must be appreciated before these predictive models are used in clinical practice.

This analysis and the predictions obtained from the models employed highlights three important findings. Firstly, even in the 'best' patient with the most favourable patient characteristics and a large functional improvement a highly successful operation does not guarantee the highest levels of satisfaction. Secondly, in the 'worst' patient with the most unfavourable patient characteristics and poor functional improvement a failure to produce a successful operation will almost certainly result in a dissatisfied patient. Thirdly, there is a spectrum of satisfaction after knee replacement which varies significantly dependent upon both pre- and post-operative factors.

Analysis 6: Summary

Patient satisfaction is increasingly being used as a comparative measure when assessing the quality of care following joint replacement in European healthcare systems (Danish registry-AR 2010). Surgeons must now not only produce technically proficient outcomes but these must also translate into high levels of patient satisfaction. The most important determinants of satisfaction are the patient's perception of the success of their operation and post-operative function. Pre-operative variables have a minimal influence upon post-operative

satisfaction bringing into question the appropriateness of restricting access to care based on arbitrary pre-operative thresholds.

Analysis 6: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

In contrast to the rest of the analyses undertaken as part of this thesis, the majority of this analysis (ordinal regression; structured equational modelling) was undertaken by Prof Stephen Rushton. By comparing this analysis to the others one notices appreciable differences in the analyses undertaken by myself, a clinician with an interest in statistics, and by Prof Rushton who is a biological modeler and statistician. This analysis is more statistically complex, using methods that are difficult for the average clinician to understand and interpret. How the findings might be applied to everyday clinical practice is not as clear as those presented elsewhere. The reader has to trust the analysis and its interpretation rather than being able to follow it and 'see it with their own eyes'. This highlights the need for clinical experts to be involved in the design and analysis of projects using registry data. These projects must be understandable and relevant to orthopaedic clinicians for there findings to be accepted and incorporated into everyday practice. Clinicians, and especially those with an understanding of registry research methodology, can help guide this process and must continue to be an integral part of the ongoing analysis of registry data.

Due to the complexity of the structured equational modelling we were only able to consider a small number of variables within these preliminary models. The addition of additional variables, such as those identified with the structured literature review, may have improved the model further. We had difficulty modelling with the OKS and EQ5D and it may be that the inclusion of other general health metrics which were more sensitive might have been

better suited to these types of analytical techniques. The inclusion of additional / alternative variables may have yielded different results. I do not, however, think it would have changed the major finding of this analysis, namely that post-operative factors such as the success of surgery and the final health outcomes were much more important than pre-operative 'demographic' factors in determining the level of post-operative satisfaction.

Unfortunately the ordinal scales used to assess satisfaction and success within the PROMs project have not been validated. They do, however, mirror similar scales used for assessing patient reported satisfaction in national cohorts (Robertsson 2000, Bourne 2006, Noble 2010). While the content of these scales differs, their construct is similar with patients asked to rate their experience ranging from bad to good with responses presented as an ordered Likert scale. The benefit of these scales is that they give a simple representation of the patient's perceptions of the results of surgery and offer an alternative to the large number of frequently used validated outcome scores that fail to appreciate an individual patient's experiences and instead focus on hard symptomatic endpoints. There is undoubted value in measuring patient satisfaction, but work needs to be done to validate these measures and standardise reporting of this metric.

What makes a patient satisfied with their surgery is still poorly understood. In similarity to previously published work this analysis has focused on investigating how defined measurable variables influence satisfaction. However, it may well be that there are a range of unmeasured or difficult to quantify variables (e.g. how expectation is handled, rapport between patient and surgical team, level of nursing care, processes of surgical care, hospital environment, post-operative after care and support etc.) that are equally, if not more, important. Future work needs to focus on defining what makes patients satisfied and what patients mean when they say they are satisfied. This will probably require a qualitative approach to research, in contrast to the quantitative approaches used previously. Only by understanding what satisfaction is can we design validated outcome tools to assess it and then make strides to improve levels of satisfaction. The ultimate goal has to be able to

tailor our approach to surgery based on a variety of patient, hospital, surgical, environmental factors so that patients at high risk of dissatisfaction can be identified and offered additional input and support to subvert the risk of a poor outcome.

5.4 Analysis 7: Analysis of the relationship between Body Mass Index and PROMs

Analysis 7: Aims

1. To determine the nature of the relationship between pre-operative, post-operative and the improvements in PROMs and obesity.
2. To establish if there evidence to support the idea of an obesity "cut off" as a basis for withholding surgery based on functional outcomes.

Analysis 7: Methods

This analysis was performed on the NJR-PROMs dataset described in section 3.4.4. From this dataset data was extracted on all patients for whom the primary indication for TKR was osteoarthritis. In total 22,691 NJR-PROMs records were available for analysis. From this cohort a further 9,018 records were excluded because BMI data was either missing or outside the range 15-60kg/m^2. This range in line with the policy of BMI analysis used by the NJR as part of their annual reporting as it is felt BMIs outside this range may represent erroneously entered data. In total 13,673 cases fulfilled these criteria and were used as the basis for analysis. The demographic details of those patients who had their BMI recorded and those who did not are given in table 5.8. As previously discussed in analysis 5 (section 5.2) the two groups were similar in respect to their patient characteristics but differed in respect to their surgical characteristics.

	Study Cohort n=13,673	Missing BMI data n=9018
Mean BMI	31.0 kg/m^2 (S.D 5.5) (95% CI 30.9 to 31.1)	NA
Patients with BMI >30 kg/m2	7771 (57%)	NA
Mean Age	69.7 years (S.D 8.8)	69.7 years (S.D 8.9)
Gender		
- Male	6117 (45%)	3896 (43%)
- Female	7556 (55%)	5229 (57%)
ASA grade		
- 1	1424 (10%)	883 (10%)
- 2	10,077 (74%)	6981 (76%)
- 3/4	2172 (16%)	1261 (14%)
Pre-operative general health		
- Excellent	483 (4%)	322 (4%)
- Very Good	3433 (25%)	2315 (25%)
- Good	6112 (45%)	4070 (45%)
- Fair	2899 (21%)	1914 (21%)
- Poor	385 (3%)	304 (3%)
- Missing	361 (3%)	200 (2%)
Number of Co-morbidities		
- 0	4933 (36%)	3332 (36%)
- 1	5480 (40%)	3541 (39%)
- 2 or more	3260 (24%)	2252 (25%)
Mean Pre-operative OKS	18.9 (95% CI 18.8 to 19.0)	18.9 (95% CI 18.7 to 19.1)
Mean Pre-operative EQ5D index	0.389 (95% CI 0.384 to 0.394)	0.387 (95% CI 0.381 to 0.393)
Mean Pre-operative EQ5D VAS	69.0 (95% CI 68.7 to 69.3	68.9 (95% CI 68.5 to 69.3)

Table 5.8: Demographics for the Patient Reported Outcome Measures (PROMs) cohorts with and without Body Mass Index (BMI) data.

For this analysis the primary outcome variables were 1) the pre- and post-operative OKS and EQ5D index 2) the change in the OKS and EQ5D index 3) the rates of post-operative complications. To help determine the relationship between BMI and these three outcomes it was analysed as both a continuous and categorical variable. Initial analysis involved the construction of scatterplots plotting BMI against the pre-operative, post-operative and change from baseline score or each of the PROMs outcomes. Scatterplots were stratified by age (≤ and >65 years) and gender and for each plot the associated linear regression equation was calculated.

Additional statistical analysis was undertaken by categorising the data into 3 groups (15-24.9kg/m^2, 25-39.9kg/m^2, 40-60kg/m^2) based on the World Health Organization (WHO) upper cut off for normal weight (BMI 25kg/m^2) and the lower cut off for morbid obesity (BMI 40kg/m^2) (WHO 2011). To limit the confounding effect of patient variables known to influence PROMs (age, gender, American Society of Anesthesiologists (ASA) grade, number of co-morbidities and general health rating) multiple linear regression was used to adjust for differences in these variables when comparing the three BMI groups. This "adjustment" effectively accounts for any differences in the distribution of these covariates and their respective influences upon the changes in the PROMs within each of the groups. Model diagnostics were satisfactory for the regression models used.

Analysis 7: Results

The mean BMI for the 13,673 patient identified was 31.0kg/m^2 (S.D 5.5) and 7771 (57%) had a BMI of ≥30kg/m^2 (table 5.8). Scatterplots of BMI plotted against the pre-operative, post-operative and change in OKS and EQ5D index stratified by age and gender are given in figures 5.4, 5.5, 5.6 and 5.7. Initial analysis of the EQ5D health scale demonstrated patients reported only minimal improvements, which were not felt to be clinically significant, using this assessment modality (mean improvement 3.0 points (95% CI 2.7 to 3.3) on a scale 0-100). The graphical output this variable is therefore not presented.

For all PROMs a consistent trend for decreasing pre and post-operative scores as BMI increased was observed irrespective of age and gender. Overall the gradient of the linear regression line for the OKS (+0.028, p=0.08), EQ5D index (+0.001, p=0.005) and EQ5D VAS (+0.033, p=0.35) were positive indicating that whilst the pre and post-operative scores decreased with BMI the trend was for the change in score to improve, albeit minimally, as BMI increased.

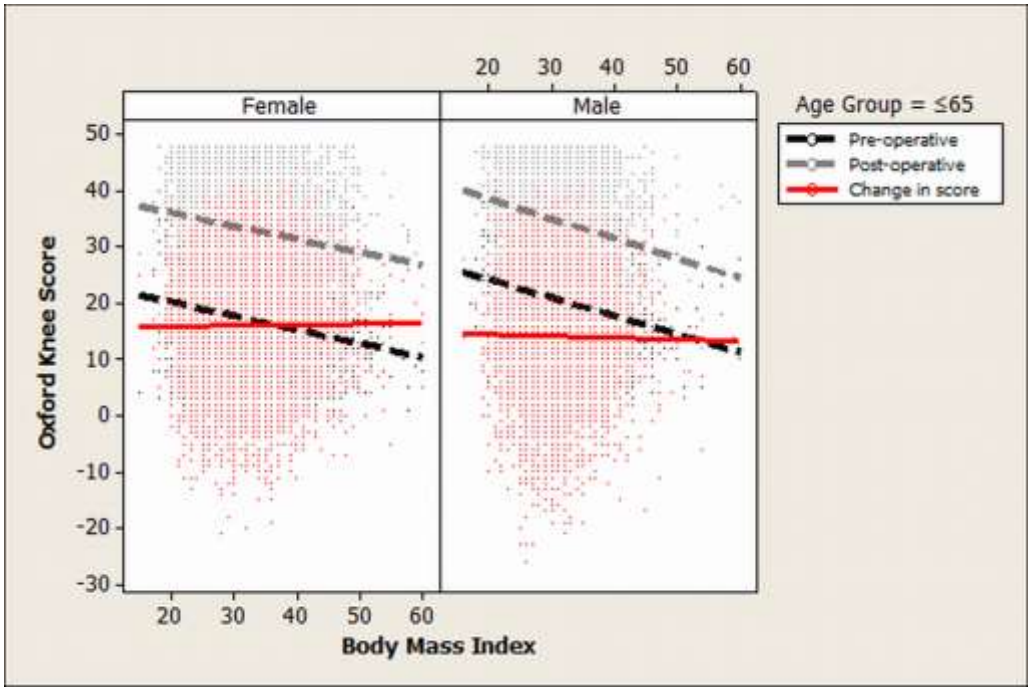


Figure 5.4: Scatterplot by gender with corresponding linear regression lines for Body Mass Index (BMI) against the pre-operative, post-operative and change from baseline for the Oxford Knee Score (OKS) in patients aged ≤65 years.

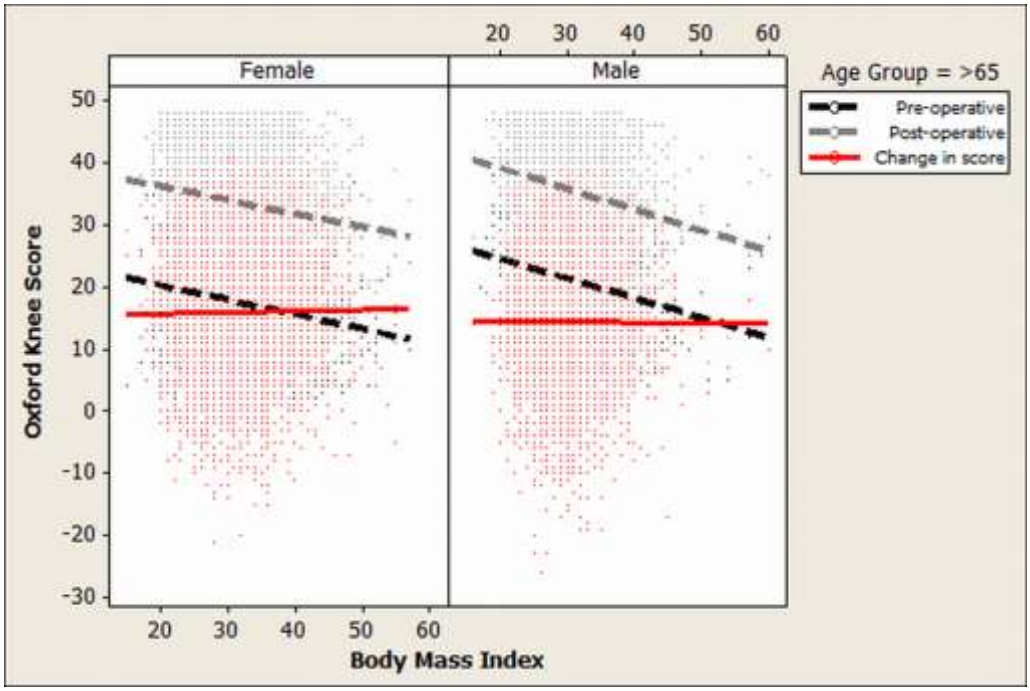


Figure 5.5: Scatterplot by gender with corresponding linear regression lines for Body Mass Index (BMI) against the pre-operative, post-operative and change from baseline for the Oxford Knee Score (OKS) in patients aged >65 years.

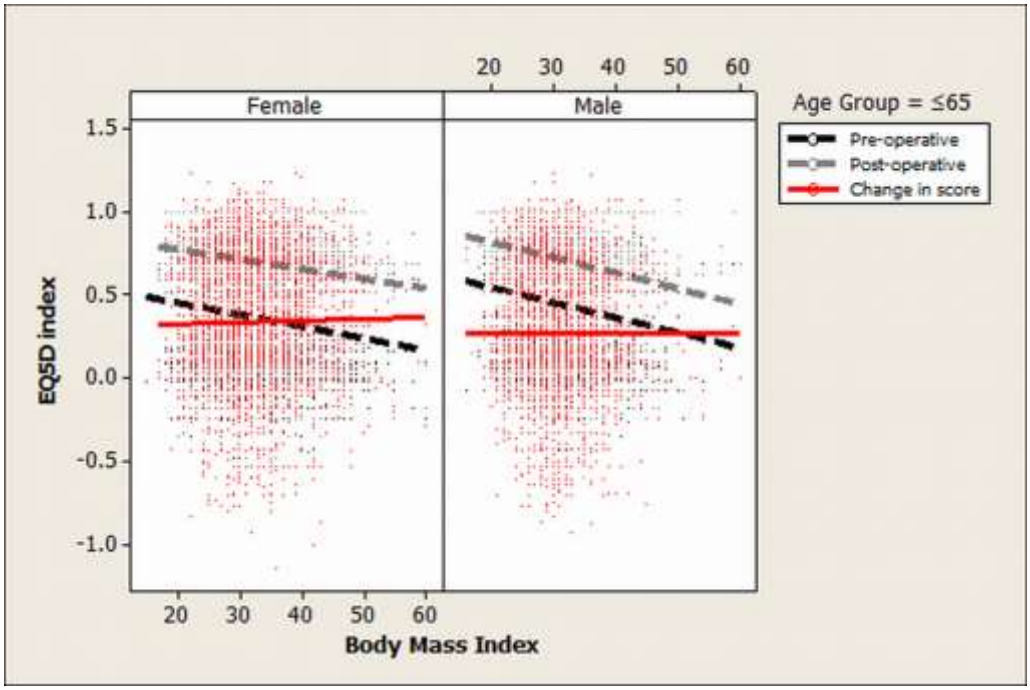


Figure 5.6: Scatterplot by gender with corresponding linear regression lines for Body Mass Index (BMI) against the pre-operative, post-operative and change from baseline for the Euroqol (EQ5D) index in patients aged ≤65 years.

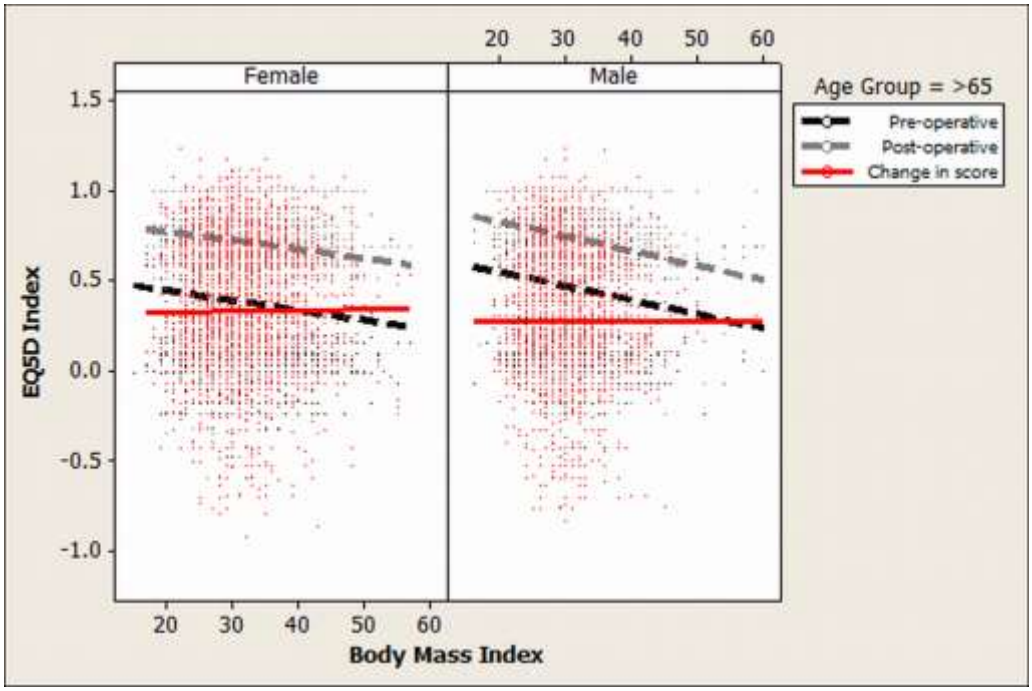


Figure 5.7: Scatterplot by gender with corresponding linear regression lines for Body Mass Index (BMI) against the pre-operative, post-operative and change from baseline for the Euroqol (EQ5D) index in patients aged >65 years.

The PROMs for the categorised BMI groups prior to adjustment are given in table 5.9. These mirrored the findings observed in figures 5.4, 5.5, 5.6 and 5.7 with the higher BMI group (40-60kg/m^2) having lower pre and post-operative OKS and EQ5D index scores but a greater change in score compared to the other two groups. The changes for the EQ5D VAS were similar for all three groups.

Unadjusted PROMs	**BMI 15-24.9kg/m^2 n = 1292**	**BMI 25-39.9kg/m^2 n = 11,363**	**BMI 40-60kg/m^2 n = 1018**	**p value (ANOVA)**
OKS				
Pre op	20.9 (95%CI 20.4 to 21.3)	19.1 (95%CI 19.0 to 19.3)	15.1 (95%CI 14.6 to 15.5)	<0.001
Post op	36.5 (95%CI 36.0 to 37.1)	34.2 (95%CI 34.0 to 34.3)	31.1 (95%CI 30.4 to 31.7)	<0.001
Change	15.7 (95%CI 15.2 to 16.2)	15.0 (95%CI 14.8 to 15.2)	16.0 (95%CI 15.4 to 16.6)	0.002
EQ5D Index				
Pre op	0.457 (95%CI 0.440 to 0.475)	0.416 (95%CI 0.410 to 0.422)	0.294 (95%CI 0.273 to 0.314)	<0.001
Post op	0.766 (95%CI 0.752 to 0.779)	0.718 (95%CI 0.713 to 0.723)	0.632 (95%CI 0.613 to 0.650)	<0.001
Change	0.309 (95%CI 0.291 to 0.327)	0.303 (95%CI 0.296 to 0.309)	0.337 (95%CI 0.316 to 0.359)	0.008
EQ5D VAS				
Pre op	71.7 (95%CI 70.6 to 72.7)	69.3 (95%CI 68.9 to 69.7)	62.7 (95%CI 0.61.4 to 64.0)	<0.001
Post op	74.7 (95%CI 73.6 to 75.8)	72.3 (95%CI 72.0 to 72.7)	65.7 (95%CI 64.4 to 67.1)	<0.001
Change	3.1 (95%CI 1.9 to 4.2)	3.0 (95%CI 2.6 to 3.4)	3.0 (95%CI 1.5 to 4.6)	0.99

Table 5.9: Analysis of the unadjusted pre-operative, post-operative and change in PROMs for the three Body Mass Index (BMI) groups.

The change in OKS, EQ5D index and EQ5D VAS for each group adjusted for age, gender, ASA grade, number of co-morbidities and general health rating are given in table 5.10. After adjustment the mean change in OKS was greatest for the BMI 40-60kg/m^2 group (mean difference vs. BMI 15-24.9kg/m^2 group 0.5 (95%CI -0.5 to 1.5), p=0.78, mean difference *versus* BMI 25-39.9kg/m^2 group 0.9 (95%CI 0.1 to 1.6), p=0.03). The mean difference

between the BMI 15-24.9kg/m^2 and BMI 25-39.9kg/m^2 groups was 0.4 (95%CI -0.3 to 1.1), p=0.57. For the adjusted EQ5D index the mean change was also greatest for the BMI 40-60kg/m^2 group although the differences with the other groups were not significant (mean difference *versus* BMI 15-24.9kg/m^2 group 0.014 (95%CI -0.021 to 0.048), p=1.00, mean difference *versus* BMI 25-39.9kg/m^2 group 0.019 (95%CI -0.008 to 0.045), p=0.29). The mean difference between the BMI 15-24.9kg/m^2 and BMI 25-39.9kg/m^2 groups was 0.010 (95%CI -0.048 to 0.021), p=1.00. The adjusted EQ5D VAS the mean change was greatest for the BMI 15-24.9kg/m^2 group although when compared to the other groups the differences were not significant (mean difference *versus* BMI 40-60kg/m^2 group 1.9 (95%CI -0.4 to 4.1), p=0.13, mean difference *versus* BMI 25-39.9kg/m^2 group 0.5 (95%CI -1.0 to 2.1), p=1.00). The mean difference between the BMI 15-24.9kg/m^2 and BMI 40-60kg/m^2 groups was 1.3 (95%CI -0.4 to 3.1), p=0.20. For all PROMs the differences observed for the changes in scores between groups were small and below the levels felt to indicate clinical significance (OKS 3 points, EQ5D index 0.1 points, EQ5D VAS 5 points).

Adjusted PROMs	**BMI 15-24.9kg/m^2 n = 1292**	**BMI 25-39.9kg/m^2 n = 11,363**	**BMI 40-60kg/m^2 n = 1018**	**p value (ANOVA)**
OKS change	15.4 (95%CI 14.9 to 16.0)	15.1 (95%CI 14.9 to 15.2)	15.9 (95%CI 15.3 to 16.5)	0.02
EQ5D Index change	0.309 (95%CI 0.291 to 0.327)	0.304 (95%CI 0.297 to 0.310)	0.323 (95%CI 0.301 to 0.344)	0.24
EQ5D VAS change	3.6 (95%CI 2.5 to 4.8)	3.1 (95%CI 2.7 to 3.5)	1.7 (95%CI 0.2 to 3.3)	0.11

Table 5.10: Analysis of the change in Patient Reported Outcome Measures (PROMs) for the three Body Mass Index (BMI) groups after adjustment using multiple linear regression.

Patient reported complications are given in table 5.11. Rates of wound complications increased as BMI increased (BMI 15-24.9kg/m^2 121 of 1292 (9%); BMI 25-39.9kg/m^2 1351 of 11,363 (12%); BMI 40-60kg/m^2 168 of 1018 (17%), p<0.001). The rates of other complications (bleeding problems, readmission, need for further surgery) were, however, similar for the three groups.

Patient reported complication	BMI 15-24.9kg/m^2 n = 1292	BMI 25-39.9kg/m^2 n = 11,363	BMI 40-60kg/m^2 n = 1018	p value
Wound Problems (%)	121 (9%)	1351 (12%)	168 (17%)	<0.001
Bleeding Problems (%)	68 (5%)	721 (6%)	66 (6%)	0.30
Readmission (%)	127 (10%)	1083 (10%)	101 (10%)	0.88
Further Surgery (%)	49 (4%)	394 (4%)	25 (2%)	0.18

Table 5.11: Patient reported complications for the three Body Mass Index (BMI) groups.

Analysis 7: Discussion.

This analysis demonstrates that although patients with high BMIs had lower post-operative knee and general health scores, the improvements they experienced were comparable in magnitude to those of patients with a lesser BMI. Obese patients thus gained as much benefit from knee replacement as patients with a "normal" BMI, even if they do not end up at a similar post-operative level. Other than wound problems, the rates of complications were not significantly different for morbidly obese patients (BMI 40-60kg/m^2) when compared to those with a "normal" (15-24.9kg/m^2) BMI.

A recent review concluded that there was currently insufficient evidence to form a definitive view on functional and quality-of-life outcomes following TKA when comparing obese and non-obese patients (Dowsey 2008). This analysis found that BMI is related to both the pre and post-operative OKS and EQ5D index in a linear fashion, with morbidly obese patients having significantly lower scores than patients with a normal BMI. The improvements in both of these PROMs for morbidly obese patients were, however, equivalent to patients with a "normal" BMI and did not "tail off" above a certain BMI threshold. This suggests there is as much benefit, in terms of improving knee function, general health and quality of life to be gained by operating on these patients. Considering solely the post-operative scores discriminates against patients with higher BMIs by suggesting they cannot gain the success achieved by non-obese patients and ignoring the fact that their functional gains are equivalent. This brings into question the validity of the arbitrarily determined BMI

thresholds implemeted by some healthcare commisioning organisations as a barrier to surgery.

The observed trend of lower pre-operative knee scores as BMI increases also suggests their may be a selection bias against obese patients relating to the point at which surgery is offered surgery. This may reflect current referal guidelines determined by healthcare comissioners (NHS North Lincolnshire 2011, NHS Warwickshire 2011) stating patients over predefined BMI thresholds should lose weight before undergoing joint replacement, or a reluctance on the part of the operating surgeon to undertake such cases. These "delays" to surgery allow progression of the natural disease process and a worsening of symptoms. This observation may also be related to the influence obesity itself has on the functional elements of the PROMs scores, such as difficulty with mobilty or ascending and descending stairs due to patient habitus and associated comorbidites, rather than functional limitations attibutable to the knee.

Other than wound problems the rates of complications between the three BMI groups did not significanly differ. These findings contrast with a number of previous studies showing higher rates of complications in the obese population (Winiarshi 1998, Foran 2004, Namba 2005, Amin 2006, Dowsey 2010). This may reflect differences in patient and surgeon reporting of complications. Obese patients may be more likely to report wound problems if they are informed pre-operatively that they are at greater risk of these problems. This could be a potential source of bias when using the patient's reported wound problems to compare obese and non-obese patients. This study did not assess the impact of BMI upon implant survival which is inferior in morbidly obese patients with 5 year survival rates of only 74% (Amin 2006). While it is important to consider complication rates and implant survivourship when making decisions regarding surgery one must remember that it is the PROMs outcomes that are linked to improvements in quality of life. Thus the value of TKR is primarily related to these scores. If patients are fully informed of the elevated risks for equivalent improvements in PROMs then these factors should not, in themselves, be a

reason to withhold surgery. There are many factors that become increasingly prevelant as BMI increases (comorbidities, technical feasibilty, anaesthetic issues), each of which may preclude surgery. BMI, however, should not in itself be a barrier to surgery. The decision to proceed should lie with the treating orthopaedic surgeon after careful consideration of all aspects of the case and the likely impact of surgery at an individual level.

Analysis 7: Summary

Although increasing BMI is associated with poorer pre and post-operative PROMs, the improvement experienced by patients was similar irrespective of BMI. It is crucial that those in control of health resources do not penalise obese patients based on the limited functional outcome data available from previous studies and a selective view of post-operative scores alone, in which they are disadvantaged by lower pre-operative scores. Instead the overall improvements in function and general health should be the barometer of success.

Analysis 7: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

Body Mass Index is poorly recorded within the NJR database. In the early years of the registry (2004/2005) less than 25% of all records contained information relating to BMI. This has steadily increased to reach a figure of 55% in 2012 (NJR-AR 2012), but it is still only a fraction of all NJR records. This meant that BMI data was only available for 13,673 of the 22,691 (60%) NJR-PROMs records analysed. While the demographic details of those patients with and without BMI data were similar there were marked differences in the surgical data recorded for these two groups (see analysis 5). This suggests there may be bias in the way BMI is recorded. How this might have influenced this analysis is unclear.

We chose to remove records with missing BMI data (case deletion). Consideration was given to imputation of BMI data for the missing records to increase the number of records available for analysis. But due to the large amount of missing data and our inexperience with these techniques we did not pursue this. For future registry analyses, and if this work was to be repeated, I would strongly consider using imputation to help increase the yield of the dataset, especially for variables such as BMI for which a large volume of data was missing. Using imputation may also help to limit any confounding effects that might be associated with the exclusion of data.

The adjustment models for this analysis included only age, gender, American Society of Anesthesiologists (ASA) grade, number of co-morbidities and general health rating. In previous analyses (analysis 1 and 5) the pre-operative PROMs scores were found to be strong predictors of the improvements observed with surgery, however we chose not to include them in the adjustment of this BMI data. This was because we felt its inclusion would not reflect the situation seen in clinical practice in which obese patients present with poorer pre-operative function than their non-obese counterparts. By including the pre-operative PROMs we would have been analysing the influence BMI had on the improvements in PROMs assuming that these patients start off from a similar starting point i.e. answering the question 'how much would an obese patient improve if they started at the same level as a non-obese patient?'. We felt that the unadjusted analysis presented in table 5.9 better reflected the clinical problem faced by clinicians and was most helpful when counselling patients about the expected improvement following knee replacement. The reality is that, for whatever reason, obese patients have poorer scores pre-operatively. Because of this they have more scope for improvement and this is probably one of the reasons why they achieve equivalent levels of functional improvement to non-obese patients. Based on the information in table 5.9 clinicians can confidently tell obese patients that they will improve following surgery and that this improvement will be similar to that seen in patients who aren't obese but that they are unlikely to achieve the same functional endpoint due to that fact that they start off from a worse position.

Chapter 6: Clinical outcomes after revision knee replacement

6.1. Revision knee replacement

Currently over 650,000 total knee replacements are performed annually in the United States (Carr 2012) and over 80,000 in England and Wales (NJR-AR 2012). This number is expected to increase further over the next 20 years, as the population of the world ages and medical interventions continue to increase average life expectancy (Kurtz 2007). The exponential increase in the use of this health technology brings with it the growing spectre of the failed knee replacement. Revision knee arthroplasty is a complex and demanding procedure requiring meticulous planning, a skilled surgical team and familiarity with an adequate inventory of a variety of surgical implants. Primary TKR is a successful operation that consistently relieves pain, improves knee function and produces levels of patient satisfaction in excess of 80% (Hawker 1998, Robertsson 2000, Baker 2007). Revision surgery is less predictable, with lower rates of survival, increased rates of complications, varied patient satisfaction, and an inconsistent ability to restore or improve quality of life (Saleh 2002, Deehan 2006, Ghomwari 2009, Hossein 2010).

Revision knee replacement encompasses a heterogeneous mix of pathology and indication and is, more often than not, handled by the occasional surgeon. This makes it difficult for the surgeon, the surgical team and the rehabilitation services involved to acquire the necessary knowledge and skills required to treat these complex cases. It also magnifies the logistical difficulties for potential researchers wanting to design meaningful research studies assessing and comparing the effectiveness of these procedures. Whilst most arthroplasty registers are designed to detect revision as their ultimate end-point the information in respect to revision available from these registers is extremely limited. Mode of failure data is inconsistently collected and the majority of registries offer no insight into functional performance, especially after revision surgery. In addition there are currently no prospective clinical trials being undertaken evaluating different approaches to revision knee surgery in England and Wales (NIHR 2012).

The orthopaedic community must therefore base surgical decisions about revision knee surgery on a restricted pool of observational, often retrospective, case series data. These papers often lack the power and design to ask questions pertinent to clinical practice. As such we currently have limited information about who needs an operation?, which operation should be done, by whom and how? What implant should we use, how should it be fixed and what level of constraint is required? How are functional outcomes influenced by the indication for revision, the type of surgery performed, and the number of previous revisions (law of diminishing returns)?.

A number of previous studies have attempted to quantify the functional improvements seen with revision knee procedures by comparing them to the improvements seen after primary surgery. These comparisons have demonstrated lower rates of satisfaction following revision surgery (Revision 73% vs. Primary 86%) (Griedanus 2011). Post-operative general health measures (Short Form 12 (SF-12)) and knee scores (Hospital for Special Surgery (HSS), Oxford Knee Score (OKS) and Western Ontario and McMaster Universities Arthritis Index (WOMAC) are also lower after revision (Hanssen 1998, Hartley 2002, Saleh 2002, Griedanus 2011). These effects are however, confounded by the substantially lower pre-operative scores observed with revision (Griedanus 2011). If instead we consider the improvements in these scores from their pre-operative baseline then differences between primary and revision surgery are less consistent (Hartley 2002, Griedanus 2011). Reported improvements for the WOMAC are greater for primary TKR (Hartley 2002, Griedanus 2011) but equivalent results are seen for the OKS (Griedanus 2011) and SF-12 (Hartley 2002, Griedanus 2011). This is important as the value of an intervention is related to its ability to improve function and quality of life rather than its final end point.

One problem with these previous analyses is their inability to sub-stratify based upon the reason for revision. This is important as knee replacements are revised for a variety of reasons each of which might potentially influence outcome. The results of a revision for one reason may not be comparable with the results of a revision performed for an alternative

reason. Stratification by reason for revision may therefore help to identify groups that respond poorly to surgery and help identify the best surgical processes for treating specific patient groups.

The simplest way of stratifying knee revisions is in to two groups, septic (infected) and aseptic (non-infected) revisions. For the septic cohort two stage revision is the established gold standard (Antti-Poika 1990, Lentino 2003, Leone 2010). Single stage revision remains the domain of the surgical enthusiast and is advocated on the basis several small case series (Göksan 1992, Scott 1993, Buechel 2004, Bauer 2006, Whiteside 2011). In selected cases single stage revision may be potentially advantageous, avoiding some of the problems of two stage revision such as stiffness and arthofibrosis resulting from a period with a spacer *in situ*, and sparing the patient a second procedure (Parkinson 2011). However, despite these perceived advantages there are currently no prospective comparative studies examining the influence of these differing approaches to surgery upon functional outcome. The aseptic cohort encompasses a myriad of reasons for revision, many of which can co-exist. It can often be difficult for the treating surgeon to ascribe a single reason for failure making it difficult to perform robust analyses on this group of patients. In those cases for which a specific mode of failure can be determined it would be useful to understand how patient characteristics and surgical factors influence the functional recovery following revision, and whether they are similar to those that influence recovery after primary knee replacement. In this way patient's who derive the most / least benefit from surgery could be identified, allowing surgery to be targeted accordingly.

The identification of patients who benefit the most following revision surgery is likely to become increasingly important. Revision surgery is expensive, and as the number of revision procedures increase so too will the economic necessity to identify the patients who will benefit the most from the existing surgical technology. It is therefore important to understand how patient and surgical factors influence the outcomes of revision surgery and to understand whether certain surgical indications and specific patient characteristics

preclude expensive revision surgery. Further research is needed to determine how the outcome of revision surgery can be maximsed, especially for the groups of patients who fail to improve following surgery. Could the surgical volume of the treating centre; type of infecting organism; level of implant constraint be important?.

Increasing surgeon and centre operative volumes have been shown to improve outcome following both primary and revision knee surgery (Marlow 2010). This is consistent with the findings presented in analysis 4 (section 4.5) for unicondylar knee replacement. Increasing operative volume intuitively leads to greater familiarity and experience. This is particularly important for those surgeons and centres undertaking complex procedures, such as revision knee replacement, that are often performed infrequently (Lavernia 1995, Shervin 2007, Bozic 2010). The association between volume and outcome has recently led the paediatric cardiac surgery services within England and Wales to restructure their services. By centralising surgical expertise in fewer larger centres they aim to produce better patient outcomes, and ensure vital services are safe and sustainable for the future (NHS 2011). Benefits of this type of service model include delivery of a trained workforce of experts producing better training for surgeons in surgical centres at the forefront of modern working practices and technologies (NHS 2011). These specialist centres could ultimately form a network facilitating collaboration in research, clinical development and the sharing of knowledge (NHS 2011). This would help with future attempts to perform prospective research. Restructuring the revision service in a similar way into fewer higher volume centres could therefore be theoretically beneficial, particularly if an audit of current practice showed there were a large number of centres performing relatively few revision procedures.

To try and address some of the issues discussed above we used the NJR and PROMs data available in an attempt to answer the following questions:

1. How do the pre-operative, post-operative and change in knee specific and general health scores differ between primary and revision TKR?
2. How do the rates of patient satisfaction following primary and revision TKR differ?
3. How does the reason for revision influence the patient reported outcome following revision TKR?
4. Which patient / surgical factors influence improvements in PROMs following revision TKR?
5. How does the approach to surgery (one *versus* two stage) influence the functional outcomes and rates of complications following revision of the infected knee replacement?
6. How many centres in England and Wales are performing revision knee procedures and what volumes of these procedures are they undertaking?

These questions are addressed in the subsequent 3 analyses.

1. Analysis of the influence of reason for revision upon early PROMs following aseptic revision knee replacement
2. Analysis of functional outcomes after one and two stage revision for the infected knee replacement.
3. Analysis of centre operative volumes for revision knee replacements performed in England and Wales

6.2 Analysis 8: Analysis of the influence of reason for revision upon early PROMs following aseptic revision knee replacement

Analysis 8: Aims

To determine:

1. How the pre-operative, post-operative and change in knee specific and general health scores differ between primary and revision TKR?
2. How the rates of patient satisfaction following primary and revision TKR differ?
3. How the reason for revision influences the patient reported outcome following revision TKR?
4. Which patient / surgical factors influence improvements in PROMs following aseptic revision TKR?

Analysis 8: Methods

This analysis was performed on the NJR-PROMs dataset described in section 3.4.4. From this dataset data was extracted for all revision knee replacements (n=996) (figure 3.5). As this analysis focused on aseptic revision surgery all cases for which the stated reason for revision included 'infection' (n=195) were excluded. The nature of the PROMs data collection means that preoperative data for two-stage revisions corresponds to their function prior to the second stage procedure, not prior to their revision. For this reason we also excluded all aseptic revisions performed using a two-stage technique (n=4). The revision cohort therefore included only single-stage revisions performed for an indication other than infection (n=797). To assess the effectiveness of revision knee replacement the PROMs for these 797 revisions were compared to the PROMs for all 23,393 of the primary TKRs recorded on the NJR-PROMS database.

Patient demographics for the primary and revision TKR groups are given in table 6.1. When compared to primary TKR patients the revision patients were younger (67.8 *versus* 69.6 years), had a higher proportion of patients graded as ASA Grade 3 or 4 (21% *versus* 15%) and had a greater proportion of patients whose surgery was performed within the National Health Service (98% *versus* 90%) by a consultant grade surgeon (92% *versus* 74%).

	TKR n=23,393	**Revision TKR n=797**	**p value**
Mean Age	69.6 (S.D 9.0)	67.8 (S.D 10.0)	<0.001
Gender (%) - Female - Male	 13,223 (57%) 10,170 (44%)	 420 (53%) 377 (47%)	 =0.03
ASA (%) - 1 - 2 - 3 - 4	 2362 (10%) 17,445 (75%) 3512 (15%) 74 (0%)	 65 (8%) 562 (71%) 170 (21%) 0 (0%)	 <0.001
Number of operations performed in the NHS (%)	20,932 (90%)	780 (98%)	<0.001
Number of operations performed by a consultant surgeon (%)	17,371 (74%)	730 (92%)	<0.001
Number of patients with 3 or more comorbidities (%)	1979 (9%)	71 (9%)	0.66
Mean Follow up	209 days	209 days	

Table 6.1: Demographics for the TKR and Revision TKR groups.

PROMs examined for this analysis included the OKS, EQ5D index, and patient satisfaction. Patients recording either an “excellent”, “very good” or “good” response to the satisfaction questionnaire were classified as satisfied and those responding “fair” or “poor” as unsatisfied. For the revision TKR group we were interested in the effects of the reason for revision upon patient outcomes. Information for this variable is available within the NJR database. To overcome the problem of multiple reasons for revision on the NJR data collection form we employed a hierarchical strategy for determining the primary reason for revision. This is described in section 3.5.2. Statistical analysis was performed as described in section 3.6.

We were also interested in determining which patient and surgical factors influenced the improvements in PROMs observed after revision surgery. Preliminary analysis established that there was significant variation in the PROMs improvements dependent upon the reason for revision which might mask the influence of other factors. We therefore chose to focus on those revisions performed for aseptic loosening / lysis as this represented the largest sub-group (n=367) with the largest PROMs improvements. The demographic and surgical variables available for our regression models were similar to those used in analysis 5 (table 3.8, section 3.4.4).

The improvement in the OKS and EQ5D were used as the primary response outcomes and the standard modeling process described in section 3.6.3 was used for this analysis. Initial analysis to establish the relationship between the individual explanatory variables and magnitude of the OKS/EQ5D improvements was undertaken using one-way analysis of variance (ANOVA) and univariable regression. Multiple linear regression was then employed as described in analysis 5. Due to the smaller numbers available for this analysis a p value of $p<0.05$ was used to indicate statistical significance within each of the models. For the final models the model estimates with 95% confidence limits are provided to allow comparison of the adjusted effect size between variables. Estimates effectively represent the predicted changes in the OKS and EQ5D for that variable once the effect of all other variables included in the model are considered. For continuous variables (age, OKS and EQ5D scores) these estimates relate to the expected changes in the response outcome (OKS/EQ5D) for a unit change in the explanatory variable. For categorical variables the estimates are given relative to the base reference category. Model residuals, 2-way interactions and other checks of model adequacy were satisfactory.

Analysis 8: Results

The mean pre-operative, post-operative and overall changes in scores were greater for the primary group when compared to the revision group for both the OKS and EQ5D. The mean improvement in the OKS was 15.1 after primary surgery compared to 10.4 after revision ($p<0.001$). The relative improvement in OKS for revision TKR was therefore 69% (95%CI 64% to 74%) of the improvement observed for primary TKR. The corresponding values for the mean post-operative OKS were 34.0 and 26.6 respectively ($p<0.001$). Overall only 7% of patients failed to demonstrate an improvement using the OKS following primary surgery compared to 17% following revision surgery ($p<0.001$) (table 6.2).

The mean improvement in the EQ5D was 0.303 after primary surgery compared to 0.231 after revision ($p<0.001$). The relative improvement in EQ5D for revision TKR was therefore 76% (95% CI, 71% to 81%) of the improvement observed for primary TKR. The corresponding values for the mean post-operative OKS were 0.710 and 0.541 respectively ($p<0.001$). Overall 21% of patients failed to demonstrate an improvement using the EQ5D following primary surgery compared to 34% following revision surgery ($p<0.001$) (table 6.2). Both groups demonstrated improvements in each of the 5 EQ5D domains postoperatively; however, the improvements were more noticeable in the primary TKA group, with a greater proportion of patients ending up in Level 1 following surgery (table 6.3).

A greater proportion of patients described their surgery as “excellent” following primary TKR (5,124 of 22,960 respondents (22%)) when compared to revision TKR (102 of 786 respondents (13%)). The proportion of patients with “poor” results was greater for the revision group (Revision, 82 of 786 respondents (10%) *versus* Primary, 856 of 22,960 respondents (4%)) (figure 6.1). Overall 83% of primary TKA were satisfied compared to 66% for revisions.

	TKR	Revision TKR	p value
OKS			
Mean Pre op	18.9 (95% CI 18.8 to 19.0)	16.2 (95% CI 15.6 to 16.8)	<0.001
Mean Change	15.1 (95% CI 15.0 to 15.3)	10.4 (95% CI 9.7 to 11.1)	<0.001
Mean Post op	34.0 (95% CI 33.9 to 34.2)	26.6 (95% CI 25.8 to 27.4)	<0.001
% No improvement	7%	17%	<0.001
EQ5D Index			
Mean Pre op	0.407 (95% CI 0.4.03 to 0.411)	0.310 (95%CI 0.286 to 0.334)	<0.001
Mean Change	0.303 (95% CI 0.298 to 0.307)	0.231 (95% CI 0.205 to 0.258)	<0.001
Mean Post op	0.710 (95% CI 0.707 to 0.714)	0.541 (95% CI 0.518 to 0.565)	<0.001
% No improvement	21%	34%	<0.001

Table 6.2: Comparison of the pre-, post-operative and change in score alongside the proportion of patients with no reported improvement for the OKS and EQ5D index for the TKR and Revision TKR groups.

EQ5D dimension		**TKR**			**Revision TKR**		
		Pre op	Post op	Net change	Pre op	Post op	Net Change
Mobility	Level 1	6.0	47.6	+41.6	3.9	22.9	+19.0
	Level 2	93.8	52.2	-41.6	95.3	76.9	-18.4
	Level 3	0.2	0.1	-0.1	0.8	0.1	-0.7
Self-Care	Level 1	69.4	78.7	+9.3	57.5	62.1	+4.6
	Level 2	30.0	20.8	-9.2	41.2	37.1	-4.1
	Level 3	0.6	0.6	0.0	1.3	0.8	-0.5
Usual Activities	Level 1	8.7	42.5	+33.8	5.8	20.5	+14.7
	Level 2	78.0	52.7	-25.3	73.3	66.8	-6.5
	Level 3	13.3	4.8	-8.5	20.9	12.7	-8.2
Pain / Discomfort	Level 1	0.9	32.6	+31.7	1.4	15.4	+14.0
	Level 2	59.7	61.5	+1.8	48.2	68.1	+19.9
	Level 3	39.4	5.9	-33.5	50.3	16.5	-33.8
Anxiety / Depression	Level 1	63.0	76.7	+13.7	52.5	60.5	+8.0
	Level 2	33.5	21.1	-12.4	40.4	33.4	-7.0
	Level 3	3.5	2.2	-1.3	7.1	6.2	-0.9

Table 6.3: Euroqol domains and percentage change pre- to post-operatively for the TKR and Revision TKR groups.

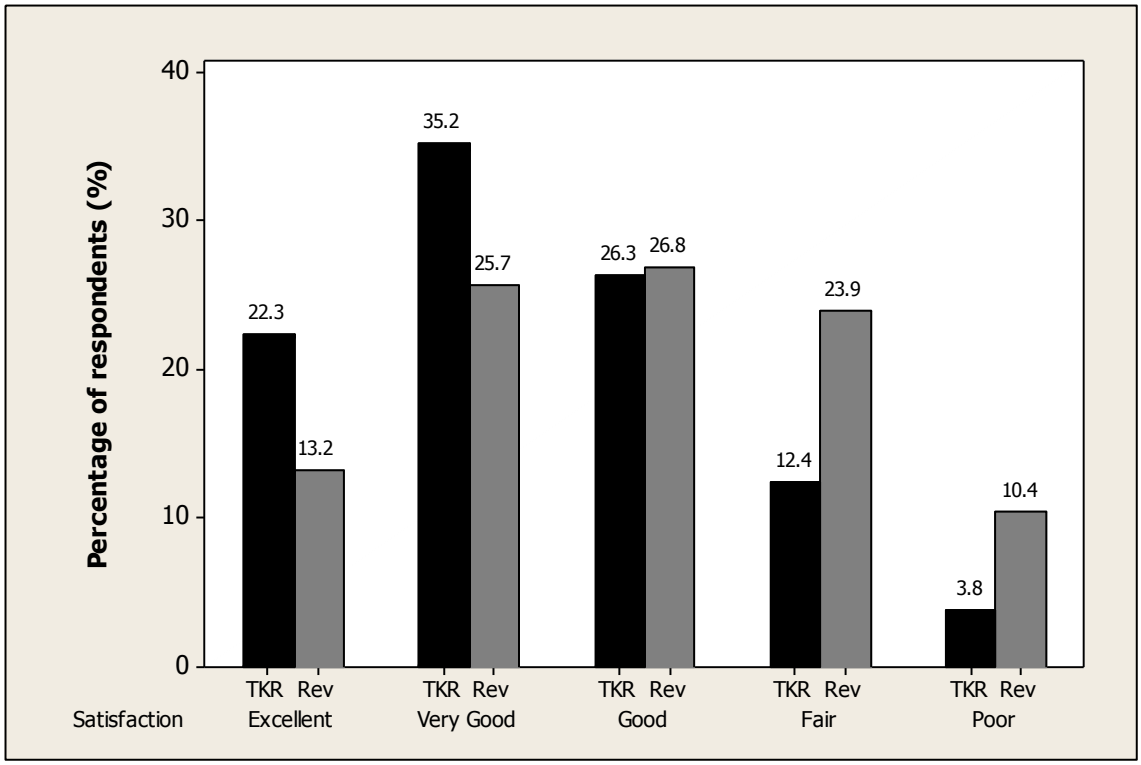


Figure 6.1: Satisfaction rating for the TKR and Revision TKR groups.

Reason for revision was seen to influence the OKS, EQ5D and satisfaction rate. The best post-operative scores were seen in patients revised for either aseptic loosening or lysis (OKS 27.8, EQ5D 0.560). Even in this group, however, the improvements in the OKS/EQ5D, the post-operative OKS/EQ5D and the rate of satisfaction were still lower than the equivalent scores following primary TKR ($p<0.001$) (table 6.4). The worst OKS and EQ5D scores were seen in revisions performed for stiffness. In this group the improvement in OKS (5.6) and the post-operative OKS (21.1) were significantly worse than the corresponding values for revisions undertaken for aseptic loosening (OKS improvement = 11.3, post-operative OKS = 27.8) ($p<0.001$). Satisfaction rates were also considerably lower between these two groups (Aseptic loosening / Lysis = 72% *versus* Stiffness = 47%, $p<0.001$).

	Malalignment	Aseptic Loosening / Lysis	Component Wear	Dislocation / Instability	Stiffness	Unexplained Pain	Other / No reason stated
N	94	367	34	81	33	84	104
Mean Age (Years)	68.6	69.3	67.8	66.0	63.9	65.5	65.9
Gender (M:F)	44:50	190:177	13:21	29:52	19:14	43:41	39:65
OKS							
Mean Pre-op	15.4 (95%CI 13.7 to 17.0)	16.5 (95% CI 15.7 to 17.3)	14.6 (95%CI 12.4 to 16.9)	17.0 (95%CI 14.9 to 19.0)	15.5 (95%CI 12.9 to 18.2)	16.8 (95%CI 15.0 to 18.5)	15.6 (95%CI 14.1 to 17.0)
Mean Change	11.3 *(95%CI 9.4 to 13.1)*	11.3 (95%CI 10.2 to 12.4)	11.6 (95%CI 8.1 to 15.2)	7.2 (95%CI 4.9 to 9.5)	5.6 (95%CI 2.0 to 9.1)	9.6 (95%CI 7.4 to 11.7)	10.5 (95%CI 8.5 to 12.6)
Mean Post op	26.7 *(95%CI 24.3 to 29.0)*	27.8 *(95%CI 26.6 to 28.9)*	26.3 *(95%CI 22.5 to 30.0)*	24.2 *(95%CI 22.0 to 26.5)*	21.1 *(95%CI 16.8 to 25.4)*	26.4 *(95%CI 23.5 to 29.3)*	26.1 *(95%CI 23.9 to 28.3)*
% No improvement	11%	16%	21%	24%	28%	15%	19%
EQ5D index							
Mean Pre-op	0.255 (95%CI 0.187 to 0.323)	0.328 (95% CI 0.292 to 0.363)	0.329 (95%CI 0.216 to 0.442)	0.329 (95%CI 0.256 to 0.403)	0.303 (95%CI 0.173 to 0.432)	0.310 (95%CI 0.232 to 0.387)	0.277 (95%CI 0.213 to 0.341)
Mean Change	0.293 *(95%CI 0.221 to 0.365)*	0.232 (95%CI 0.192 to 0.271)	0.203 (95%CI 0.069 to 0.337)	0.208 (95%CI 0.129 to 0.287)	0.176 (95%CI 0.015 to 0.338)	0.172 (95%CI 0.091 to 0.252)	0.263 (95%CI 0.191 to 0.334)
Mean Post op	0.548 (95%CI 0.479 to 0.617)	0.560 (95%CI 0.526 to 0.593)	0.533 (95%CI 0.439 to 0.627)	0.537 (95%CI 0.467 to 0.608)	0.479 (95%CI 0.334 to 0.624)	0.481 (95%CI 0.395 to 0.568)	0.540 (95%CI 0.478 to 0.601)
% No improvement	33%	33%	33%	40%	44%	35%	34%
Satisfied (%)	64%	72%	73%	60%	47%	58%	59%

Table 6.4: PROMs for the revision TKR group dependent upon reason for revision.

The only variables observed to influence improvement in the OKS and EQ5D within the initial univariate analysis of the aseptic loosening / lysis group (n=367) were age (OKS only), the pre-operative OKS (both OKS and EQ5D), the pre-operative EQ5D (both OKS and EQ5D), pre-operative general health (OKS only), and anxiety level (EQ5D only) (table 6.5). In similarity to the findings of analyses 1 (section 4.2) and 5 (section 5.2) greater PROMs improvements were observed with increasing age, a deceasing pre-operative OKS / EQ5D score, better pre-operative general health and a lower level of pre-operative anxiety.

Predictor variable	Dependent variable	
	OKS improvement	EQ5D improvement
Patient factors		
Age (years)	**0.008**	0.81
Pre-operative OKS	**<0.001**	**<0.001**
Pre-operative EQ5D index	**0.01**	**<0.001**
Number of comorbidities	0.64	0.20
Gender	0.43	0.58
Pre-operative disability	0.45	0.23
Pre-operative general health	**0.05**	0.57
Depression	0.06	0.62
Anxiety level	0.37	**0.01**
ASA grade	0.13	0.18
Surgical factors		
Lead surgeon grade	0.25	0.30
Hospital type	0.78	0.84

Table 6.5: Summary of the p values for univariable analysis with improvement in OKS and EQ5D index as the response variables.

Once the influence of other variables was considered the only factors seen to influence outcome within the multiple linear regression models were age, pre-operative OKS and pre-operative general health for the OKS (all $p<0.05$) and pre-operative EQ5D, pre-operative OKS and anxiety level for the EQ5D (all $p<0.001$) (table 6.6). The variables present within the final models and the sizes and direction of their co-efficients were similar to those observed for the primary TKR analysis performed in analysis 5 (section 5.2).

	OKS improvement			EQ5D improvement		
Predictor Variable	**Estimate**	**95% CI**	**p value**	**Estimate**	**95% CI**	**p value**
Patient variables						
Age	0.14	0.03 to 0.24	=0.001	-	-	-
Pre-operative OKS	-0.42	-0.55 to -0.29	<0.001	0.007	0.001 to 0.013	<0.001
Pre-operative EQ5D index	-	-	-	-0.922	-1.003 to -0.840	<0.001
Pre-operative general health						
Excellent	Reference			-	-	-
Very Good	1.42	-0.91 to 3.75	0.22	-	-	-
Good	2.01	0.07 to 3.95	0.04	-	-	-
Fair	-0.66	-2.84 to 1.51	0.55	-	-	-
Poor	-6.34	-10.26 to -2.42	0.002	-	-	-
Anxiety level						
No anxiety/depression	-	-	-	Reference		
Moderate anxiety/depression	-	-	-	0.052	-0.004 to 0.110	0.07
Severe anxiety/depression	-	-	-	-0.183	-0.278 to -0.088	<0.001

Table 6.6 Summary of the variables significant in the final multiple linear regression models. Estimators with 95% CI are presented to allow comparison of the effect size of each variable. For categorical variables the estimators are given as differences relative to the first category. For continuous variables the estimate describes the effect on the change in score if the predictor variable was increased by one point.

Analysis 8: Discussion

This analysis found that post-operative patient reported knee specific and general health scores and their associated improvements from baseline were greater following primary TKR when compared to aseptic revision TKR. Revision TKR has been shown to produce improvements in knee function using a variety of assessment modalities including the OKS, KSS and WOMAC (Hanssen 1998, Hartley 2002, Deehan 2006, Ghomwari 2009, Hossein 2010, Greidanus 2011). The findings from this study are comparable with results from the New Zealand registry where reported 6 month post-operative OKS scores were 37.2 and 29.4 following primary and revision TKR respectively (NJR-AR 2010). Post-operative knee scores assessed using the Hospital for Special Surgery Score (Good to Excellent results: Primary 92% *versus* Revision 81%) (Hanssen 1998), WOMAC (Primary 80.2 *versus* Revision 69.1) (Greidanus 2011) and OKS (Converted to 0-100 range: Primary 78.3 *versus* Revision 68.4) (Greidanus 2011) are consistently better for primary when compared to revision TKR. Similarly post-operative SF-12 scores have been shown to be better following primary TKR (Primary 83.5 *versus* Revision 71.6) (Hartley 2002, Greidanus 2011).

The evidence relating to differences between primary and revision with respect to the changes from baseline for these assessment tools is however lacking. Direct comparisons of primary and revision TKR demonstrated the improvements for the SF-12 (Hartley 2002, Greidanus 2011) and OKS (Greidanus 2011) were equivalent and the scores for the overall WOMAC, as well as its pain and function components were only marginally better following primary TKR (Hartley 2002, Greidanus 2011). This analysis is therefore the first to demonstrate that the improvements from baseline are smaller for patients undergoing revision procedures irrespective of assessment modality.

Rates of satisfaction following primary TKR range from 81 to 86% (Robertsson 2000, Baker 2007, Bourne 2010). Reports of satisfaction following revision vary between 73% and 88% (Hossein 2010, Greidanus 2011). The satisfaction rate for primary TKA (83%) observed in this analysis is therefore comparable. However, the rate following aseptic revision surgery (66%) is lower than previous reports. This may be a reflection of the shorter duration of follow up and the differing methods of collection for this data. Patients rated the outcome of their surgery as “Poor” in 4% of primaries and 10% of revisions. This indicates that while the overall rates of satisfaction differed, the proportion of patients reporting the poorest results was small for both groups.

Revisions performed for aseptic loosening and lysis were associated with the greatest post-operative OKS/EQ5D scores and the highest rates of satisfaction. Revisions for malalignment and component wear produced comparable outcomes when compared to aseptic loosening / lysis. The improvements in scores for revisions undertaken for dislocation/instability, unexplained pain and stiffness were typically smaller with the worst OKS/EQ5D scores and satisfaction rates seen in the stiffness group. The reason why revisions for stiffness perform poorly may be related to poorer post-operative range of motion and function. Unfortunately due to the type of data collected we could not explore

this hypothesis in further detail. There are currently no direct comparative studies assessing the effects of reason for revision upon functional outcome and satisfaction, with studies combining results from aseptic revisions without sub stratifying those performed for aseptic loosening against other reasons. In one of the few studies to examine revisions performed for aseptic loosening Bertin *et al* (Bertin 1985) found 91% of 53 revisions reported relief of pain and 80% could walk for over 30 minutes. Reports of revisions performed for stiffness demonstrated low post-operative KSS scores and only modest improvements in scores at a mean of 43 months post-operation (KSS pain post-op = 46.9, improvement = 31.9, KSS function post-op = 58.4, improvement = 18.4) (Kim 2004). These findings indicate that surgeons should expect different improvements and final functional scores dependent upon the reason for revision and that even in the "best" aseptic loosening / lysis group the expected results are lower than those observed after primary TKR.

For the aseptic loosening / lysis group the factors influencing improvements in PROMs were similar to those previously observed for primary TKR, namely the relevant pre-operative PROMs in conjunction with pre-operative general health and / or a measure of mental health (in this case the anxiety level). Interestingly a number of the factors associated with PROMs improvements in the primary setting were not associated with improvements in the cohort of aseptic loosening / lysis patients. This may be due to the fact that the aseptic loosening / lysis group was a lot smaller (367 *versus* 23,393 patients) meaning the analysis of this group was underpowered to detect 'weak' associations with marginal size effects.

Analysis 8: Summary

The observed improvements in knee function and general health following revision TKR are only 69% to 76% of those observed for primary TKR. In addition these outcomes are considerably worse in specific groups dependent upon the reason for revision. On average all revisions improved from baseline irrespective of the reason for revision although revisions for stiffness and unexplained pain produced the smallest improvements. This

information is useful as it allows surgeons to counsel patients about the expected improvements and final functional outcomes following revision relative to the levels achieved after their primary TKR. In addition it provides surgeons with information about how these outcomes change in the context of the reason for revision. Within the group of patients undergoing revision for aseptic loosening / lysis the key factors influencing the PROMs improvement were the relevant pre-operative PROMs, pre-operative general health, age and anxiety level.

Analysis 8: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

The NJR-PROMs data is not suited to the analysis of revision data. For primary TKR the majority of functional improvement has occurred by 6 months and is unlikely to improve further after this time point (Johnston 2009). PROMs data collection at 6-8 months (90% of PROMs records) is therefore appropriate. However, it takes longer to reach a functional plateau after revision TKR owing to the extent of the surgery and the time required for recovery. For revision TKR this plateau phase may not be reached until at least year after surgery (Malviya 2011) meaning further improvements occurring later during postoperative recovery could not be appreciated. While comparison of PROMs assessments at 6 months is valid it may not reflect the final endpoint for the revision cohort. Interpretation of 'post-operative' and 'improvement' scores must therefore be interpreted with caution. This highlights one of the problems with using a national dataset with standardised data collection, namely that the number and timing of post-operative assessments cannot be altered. Using these databases to assess conditions for which they were not primarily designed (revision rather than primary surgery) is therefore problematic.

The PROMs project is still in its infancy and currently only collects data from a sample of all primary and revision TKRs performed. The number of revisions with PROMs records is approximately 10% and includes less than a thousand cases with useable PROMs data. This compares to primary TKR for which >23,000 useable records are available. The smaller number of records for revision procedures limited our ability to analyse this group, especially when the groups were sub stratified by the indication for revision. For this reason the analysis of revision data was, on the whole, restricted to descriptive rather than analytical and investigative analyses.

The NJR dataset and data collection forms are designed to monitor primary implants prospectively over time and in those cases that come to revision record specific information relating to that procedure. A number of revisions are, however, revisions of primaries performed prior to the inception of the NJR or are re-revisions of prior revisions. For these groups of patients there is often limited information recorded about their previous surgery as the NJR was not designed to collect this information. One of the key determinants of outcome after revision surgery is the number of previous revisions (NZJR-AR 2012). This is almost impossible to ascertain from the NJR dataset unless the revision is being performed on a primary procedure that has also been registered with the registry. As such we could not include this important information with our analyses. Additionally, unlike many other registries, the NJR does record information about the complexity of the revision i.e. whether the revision was a major (change of all components) or minor (change of polyethylene only) revision. It may be possible to determine this information from the implant records but this would require a detailed examination of each individuals NJR records which is likely to be time consuming and potentially uninformative.

As the PROMs project grows the number of records linked to revision cases will increase. It may therefore be possible to examine this cohort in more detail in the future. However, due to some fundamental problems, including limited follow up and failure to record key clinical

information, the analysis of this group will remain limited unless changes to the way in which data is collected can be made.

6.3 Analysis 9: Analysis of functional outcomes after one and two stage revision for the infected knee replacement

Analysis 9: Aims

1. To determine how the approach to surgery (one *versus* two stage) influences the functional outcomes and rates of complications following revision of the infected knee replacement?

Analysis 9: Methods

This analysis was performed on the NJR-PROMs dataset described in section 3.4.4. From this dataset data was extracted for all revision knee replacements (n=996) (figure 3.5). As this analysis focused on septic revision surgery only those cases for which the stated reason for revision included 'infection' (n=195) were included. The 195 revisions included 33 single stage revisions, 73 first of two stage revisions and 89 second of two stage revisions. Data for both the first and second stage of the two stage procedures was available for only 5 patients within the two stage group. Due to the limited numbers with complete data it was therefore impossible to generate meaningful change scores spanning both the first and second of the two stage procedures for this group. Analysis therefore focused on comparing final outcomes for the single and two stage group by comparing the post-operative PROMs for the single stage and second of two stage groups. This comparison required the assumption that the two groups were similar prior to surgery i.e. they started off in the same place prior to the first stage of surgery. This was felt to be a valid assumption as the pre-operative demographic variables for the single stage and first of two stage groups were similar (table 6.7). Analysis of the combined pre- and post-operative PROMS data for the single stage revision group was performed to demonstrate the effect size in the improvements seen after revision for infection for this group of patients.

PROMs examined for this analysis included the OKS, EQ5D index, patient satisfaction and rates of post-operative complications. Patients recording either an “excellent”, “very good” or “good” response to the satisfaction questionnaire were classified as satisfied and those responding “fair” or “poor” as unsatisfied. Statistical analysis was performed as described in section 3.6.

	Operation Type		
	Single Stage	2nd of 2 stage	p value
N	33	89	
Mean Age (Years)	69.4 (S.D 10.7)	70.3 (S.D 8.9)	0.67
Gender • Female • Male	 15(46%) 18(55%)	 35(39%) 54(61%)	0.54
ASA grade • 1 • 2 • 3/4	 1(3%) 22(67%) 10(30%)	 3(3%) 56(63%) 30(34%)	0.93
Hospital Type: NHS hospital Other hospital	 32 (97%) 1 (3%)	 89 (100%) 0 (0%)	-
Lead Surgeon: • Consultant • Other grade	 31 (94%) 2 (6%)	 84(94%) 5(6%)	0.93
Number of comorbidities • 2 or less • 3 or more	 28 (85%) 5 (15%)	 78 (88%) 11 (12%)	0.69

Table 6.7: Demographic details for the two infected revision groups.

Analysis 9: Results

The overall mean post-operative OKS for the single and second of two stage revisions was 23.4 (95% CI 21.2 to 25.6) and for the EQ5D index was 0.484 (95% CI 0.414 to 0.554). There were no significant differences between either group for the post-operative OKS or EQ5D index (Table 6.8). The distribution of patients in each level of the 5 EQ5D domains was also comparable (Table 6.9). The proportions of patients reporting Excellent/Very good/Good

levels of satisfaction were similar between the two groups (single = 20 of 33 respondents (61%) vs. two stage = 50 of 87 respondents (57%) (p=0.76)). Overall 8 patients (24%) from the single stage and 21 patients (24%) from the two stage groups felt the result of their operation was poor (p=0.99).

	Procedure type	Mean	S.D	95% CI	p value
OKS Score	Single stage revision	25.1	13.1	21.1 to 29.1	0.39
	2nd of 2 stage revision	22.8	12.4	20.2 to 25.4	
EQ5D Index	Single stage revision	0.498	0.397	0.364 to 0.632	0.75
	2nd of 2 stage revision	0.473	0.355	0.397 to 0.548	

Table 6.8: Comparison of the post-operative OKS and EQ5D for the single stage and second of two stage groups.

EQ5D index dimension		Single Stage	Second of two Stage
Mobility	Level 1	18%	14%
	Level 2	76%	85%
	Level 3	6%	1%
Self-Care	Level 1	58%	48%
	Level 2	36%	48%
	Level 3	6%	5%
Usual Activities	Level 1	18%	14%
	Level 2	61%	59%
	Level 3	21%	27%
Pain / Discomfort	Level 1	19%	18%
	Level 2	63%	62%
	Level 3	19%	20%
Anxiety / Depression	Level 1	58%	57%
	Level 2	33%	34%
	Level 3	9%	8%

Table 6.9: Breakdown of the post-operative EQ5D index by its 5 constituent domains for the two groups. Level 1 = no problems, Level 2 = moderate problems, Level 3 = severe problems. Numbers represent the percentage of patients at that particular level for each domain.

When comparing rate of complications there were no differences in the rates of self-reported complications between the two groups (table 6.10).

	Operation group		
Complication	Single Stage n = 33	Two Stage n = 89	p value
Post-operative bleeding	5 (15%)	14 (16%)	0.94
Wound problems	8 (24%)	30 (34%)	0.32
Requirement for further surgery	7 (21%)	23 (26%)	0.60
Requirement for readmission	8 (24%)	31 (35%)	0.27

Table 6.10: Complications recorded for the two infected revision groups.

For reasons previously outlined it was not possible to link the second stage to its corresponding first stage meaning that preoperative function of the two stage group could not be adequately assessed. To assess the size effect of the functional improvements associated with revision the paired pre and post-operative data for the subset of single stage revisions was compared. The mean pre and post-operative OKS were 15.3 (95% CI 13.0 to 17.6) and 25.1 (95% CI 21.1 to 29.1) respectively giving a mean improvement of 9.8 (95% CI 5.2 to 14.4). For the EQ5D index the equivalent pre and post-operative means were 0.254 (95% CI 0.129 to 0.379) and 0.498 (95% CI 0.364 to 0.632) with a mean improvement of 0.244 (95% CI 0.109 to 0.379). The pre and post-operative EQ5D domains scores are given in table 6.11.

EQ5D index dimension		Pre-operative	Post-operative
Mobility	Level 1	6%	18%
	Level 2	88%	76%
	Level 3	6%	6%
Self-Care	Level 1	58%	58%
	Level 2	39%	36%
	Level 3	3%	6%
Usual Activities	Level 1	3%	18%
	Level 2	52%	61%
	Level 3	46%	21%
Pain / Discomfort	Level 1	3%	19%
	Level 2	49%	63%
	Level 3	49%	19%
Anxiety / Depression	Level 1	58%	58%
	Level 2	39%	33%
	Level 3	3%	9%

Table 6.11: Breakdown of the pre and post-operative EQ5D index by its 5 constituent domains for the single stage revision group (n=33). Level 1 = no problems, Level 2 = moderate problems, Level 3 = severe problems. Numbers represent the percentage of patients at that particular level for each domain.

Analysis 9: Discussion

This analysis failed to demonstrate any differences in the final PROMs scores when two different approaches to treatment of the infected knee replacement were compared. Two stage revision is currently considered the gold standard procedure for treatment of the infected knee arthroplasty (Leone 2010). One stage revision may be successful in carefully selected patients in the hip (Moyad 2008), and has been described in the knee (Göksan 1992, Scott 1993, Buechel 2004, Bauer 2006, Whiteside 2011). Existing literature on two stage revision is plentiful, consisting mainly of retrospective case series, and suggests positive results with the exception of a few reports where outcomes have been poor due to certain factors such as virulent organisms and revision after failed debridement (Mittal 2007, Macmull 2010, Gardner 2011). The reports relating to single stage revision are more sparse and are generally of poorer quality. However, positive results, comparable to two-stage revision have been achieved in enthusiastic centres (Jämsen 2009).

Gooding *et al* reported a mean satisfaction score of 71 out of 100 in 115 retrospectively reviewed two-stage revisions for infection using a PROSTALAC spacer (Gooding 2011). No further breakdown of satisfaction data was presented within this report. The overall satisfaction (patients rated as Excellent/Very Good/Good) within this analysis was only 58%, however, this may be related to the way in which satisfaction is recorded within the PROMs and the method we employed for turning a 5 point scale into a binary outcome.

A number of studies have used validated knee scores, usually the Knee Society or Hospital for Special Surgery scores, to quantify outcome following revision for infection (Laffer 2006). Jamsen *et al* reviewed 31 studies reporting outcomes after knee revision for infection, and found that the majority of patients scored above 80 out of 100 for these assessment modalities, equating to scores in the 'excellent' category (Jämsen 2009). No difference was observed between single and two stage revision. In one of the few comparative studies to date Bauer *et al* found no difference in the clinical and functional results between one and two stage procedures in a series of 30 single stage and 77 two stage revisions (Bauer 2006). For the cohort presented here the overall mean post-operative OKS was 23. This compares to a mean OKS of 24 reported by Gooding at an average of 9 years follow up (Gooding 2011). This emphasises that revision for infection is associated with poor functional results, especially when compared the mean OKS of 35 reported after primary TKR (Analyses 1 and 5).

A number of small case series have examined functional outcomes and satisfaction after single stage revision for infection and suggest they are good (Göksan 1992, Scott 1993, Whiteside 2011). Of the 17 knees reviewed by Scott *et al* no one reported an increase in their level of pain following revision and 15 patients reported either no pain or a significant improvement in their level of pain (Scott 1993). Göksan reported severe pain on walking in just 2 of their 18 single stage revisions (Göksan 1992). Walking aids were used by 11

patients but 8 of these for reasons other than the revision knee. More recently Whiteside *et al* reported a mean Knee Society Score of 83 (out of 100) at two years follow up in their series of 18 single stage revisions (Whiteside 2011).

Due to the prospective nature of the PROMs data collection it has been possible to quantify the improvements gained following revision in a cohort of single stage patients. Improvements from baseline were observed for both the OKS and EQ5D index. There were also improvements for the 5 EQ5D domains most notably involving Mobility, Ability to perform usual activities, Pain and Anxiety. These findings help quantify the value of revision in this setting and provide some information about the magnitude of improvement relative to that observed after aseptic revisions (analysis 8).

The fact that we found no differences in the postoperative OKS, EQ5D and patient reported satisfaction for one and two stage revision for infection has a number of potential explanations. An obvious problem with this analysis was the lack of statistical power due to the small numbers in the single stage group and the possibility of type II statistical error. Alternatively it could be that patients revised for infection have a final outcome that is similar at the end of the surgical episode, irrespective of the procedure performed and it is the presence of infection rather than the mode of treatment that is the key determinant of outcome. Another possible explanation could be that single stages may be functionally superior in cases where the infection is successfully eradicated but be prone to higher rates of re-infection which are associated with poorer outcomes, the result being that their overall performance is equivalent to two stage revisions. As no information on microbiological outcomes or re-infection were available, it was not possible to sub analyse dependent upon infection severity or the eradication of infection. As such it was not possible to ascertain whether singles stages were reserved for less severe cases or whether there was a relationship between re-infection and patient outcomes, particularly whether successful eradication of infection produces a better final outcome. However despite an

inability to establish re-infection rates the results presented, which include any potential re-infections, give a good overall view of patient outcomes following these two procedures.

The clinical implications of this study are that surgeons should not recommend singe stage revision for infection to patients based on the perception that it is associated with better functional outcomes and higher levels of patient satisfaction, as results suggest both single and two-stage revision are equivalent. Furthermore patients should be counselled that the functional outcome after revision for infection is likely to be poor irrespective of the type of procedure performed.

Analysis 9: Summary

There was no difference between single stage and two stage revision as assessed by a variety of patient reported outcome measures. Surgeons should not recommend singe stage revision for infection to patients based on the perception that it is associated with better functional outcomes and higher levels of patient satisfaction until further evidence is available. Prior to revision patients should be counselled that the functional outcome after revision for infection is likely to be poor irrespective of the type of procedure performed.

Analysis 9: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

In similarity to analysis 8 this analysis also suffered from methodological problems related to the ability of NJR-PROMs data to assess outcomes after revision. Functional recovery after revision for infection takes even longer than recovery following revision for aseptic causes

(Malviya 2012), raising questions about the validity of using PROMs data to assess this group of patients. Even in the revision setting the majority of functional improvement has occurred by 6 months and scores are starting to plateau. Useful data can therefore be obtained at this point, but it is unclear whether comparisons at 6 months adequately reflect differences in PROMs once full recovery has occurred.

Details relating to the type of infection (e.g. microbiological results) were unavailable meaning we could not stratify for the infecting organism. Similarly data on reasons for selection of one procedure or the other was not available due to the retrospective registry based design. The benefit lost in not having these data is, to some extent, offset by the larger numbers of cases made available through the use of a national dataset. This would not have been possible in a single centre study where microbiological and surgical episode data would have been available. In the future it may be possible to gain information about the type of infection by linking the NJR and PROMs datasets with information held by the Health Protection Agency (HPA). This may give some useful information about the influence of the infecting organism upon outcome and may be an avenue for further work in the future.

As discussed within the methods section it was not possible to link the data on second stage procedures to the first stage, and thus pre-operative data was not available for this group. We were therefore only able to compare postoperative scores (with their inherent limitations (see above)). We have to assume pre-operative parity as we had no way of knowing whether the second of two stage and single stage groups were matched prior to their revision episode. In addition no detail on the timing of second stage procedures in reference to their first stage procedure was available, a factor which may obviously play a part in outcome. One may postulate a potential selection bias, whereby single stage procedures may be done on less severe cases, positively biasing the results of this operation type. However, as the results did not show single stage revision to be superior, this potential effect has not significantly skewed the results.

Finally the numbers included in the study could be criticised as being too small. Post-hoc power analysis based on the distribution and size of the smaller single stage group showed this study to be sufficiently powered to detect a difference of 6.7 points on the OKS and 0.2 points on the EQ5D with 80% power and $p<0.05$. While these size effects are quite large, calculation shows that approximately 165 patients would be required in each arm to detect a clinically relevant size effect of 3 points on the OKS and 0.1 points on the EQ5D. Smaller size effects would therefore not be detected within this analysis. However, in defence of this analysis it must be recognised that achieving the requisite number to adequately power an analysis of these two techniques is unrealistic. As such the best way to compare these groups in the future might be through a meta-analysis of smaller analyses as presented here.

6.4 Analysis 10: Analysis of centre operative volumes for revision knee replacements performed in England and Wales

Analysis 10: Aims

1. To determine centre operative volumes for revision knee replacement by auditing all centres performing revision knee procedures in England and Wales against two pre-defined standards linked to hospital volume:
    a. Operative volume should be greater than 10 revisions per year.
    b. More than 2.5 revisions should be performed for every 100 primary arthroplasties implanted.

Analysis 10: Methods

To perform this analysis data was extracted from the NJR database for all revision knee replacement performed between 1st July 2008 and 30th June 2010. In total 359 centres undertook at least one revision during this period. For each centre information on the volume of primary and revision knee procedures undertaken during this 2 year period was available. Additional information on the hospital name, hospital type (NHS, Independent hospital, Independent Sector Treatment Centre (ISTC)) and associated NHS Trust was collected. This information was independently verified using the hospital internet home page and the Dr Foster Hospital Guide (www.drfosterhealth.co.uk/hospital-guide/#hospsearch).

For each centre we audited two aspects of surgical volume, the annual revision volume and the proportion of revisions per 100 primary procedures. These were chosen as both influence the exposure to revision and development of revision experience. Based on the work of Lavernia *et al* (Lavernia 1995), Judge *et al* (Judge 2006) and Yasunaga *et al* (Yasunaga 2009) on the effects of surgical volume in primary and revision arthroplasty we

defined hospitals as “low volume” if they performed <20 revisions over the 2-year period (10/year). Centres performing more than 20 revisions over the two years were defined as “high volume”. These centres were further sub stratified (20-49 per 2 years, 50-99 per 2 years, >99 per 2 years) based on volume to help determine the geographical spread of the highest volumes centres. The number of revisions performed per 100 primary knee arthroplasties has not been previously described. We therefore arbitrarily assigned centres performing less than 2.5 revisions per 100 primary arthroplasties as “low ratio revisers”.

Using hospital postcodes each hospital was plotted on a map of England and Wales to demonstrate the distribution of centres undertaking revision surgery and their associated surgical volume. Postcodes were linked to grid references via the National Statistics Postcode Directory (NSPD) August 2010, available from EDINA UK Borders (http://edina.ac.uk/). Mapping was undertaken in ArcGIS 9.3 (ESRI, Redlands, California, USA).

Analysis 10: Results

During the 2-year analysis period 396 different centres performed 153,133 primary knee arthroplasties. Of these 359 (91%) performed 9,659 knee revisions, equivalent to 6.2 revisions for every 100 primary arthroplasties performed. Revision centres included 208 (58%) NHS hospitals performing 8,148 revisions, 141 (39%) independent hospitals performing 1258 revisions and 10 (3%) Independent Sector Treatment Centres (ISTC) performing 253 revisions.

The median number of revisions performed per hospital was 14 over 2 years (Range 1 to 287). There was a difference in the number of revisions performed by hospital type with NHS hospitals performing a median of 28 per 2 years (Range 1 to 287) compared to 6 per 2

years (Range 1 to 84) for independent hospitals and 22 per 2 years (Range 1 to 73) for ISTCs (NHS vs. Independent, p<0.001).

Figure 6.2 demonstrates the volume of knee revision procedures performed at each centre. In total 28 centres performed only 1 revision per 2 years (table 6.5). Two hundred and twelve (59%) centres were classified as “low volume” (<20 revisions per 2 years) accounting for 1573 (16%) revisions (table 6.5). In contrast the 12 highest volumes centres (>99 revisions per 2 years) performed 2304 (24%) revisions. In total 78 of the 208 (38%) NHS hospitals were classified as “low volume” compared to 128 of 141 (91%) of the independent hospitals and 6 of 10 (60%) of the ISTCs. The geographical distribution of centres performing revisions is shown in figure 6.3.

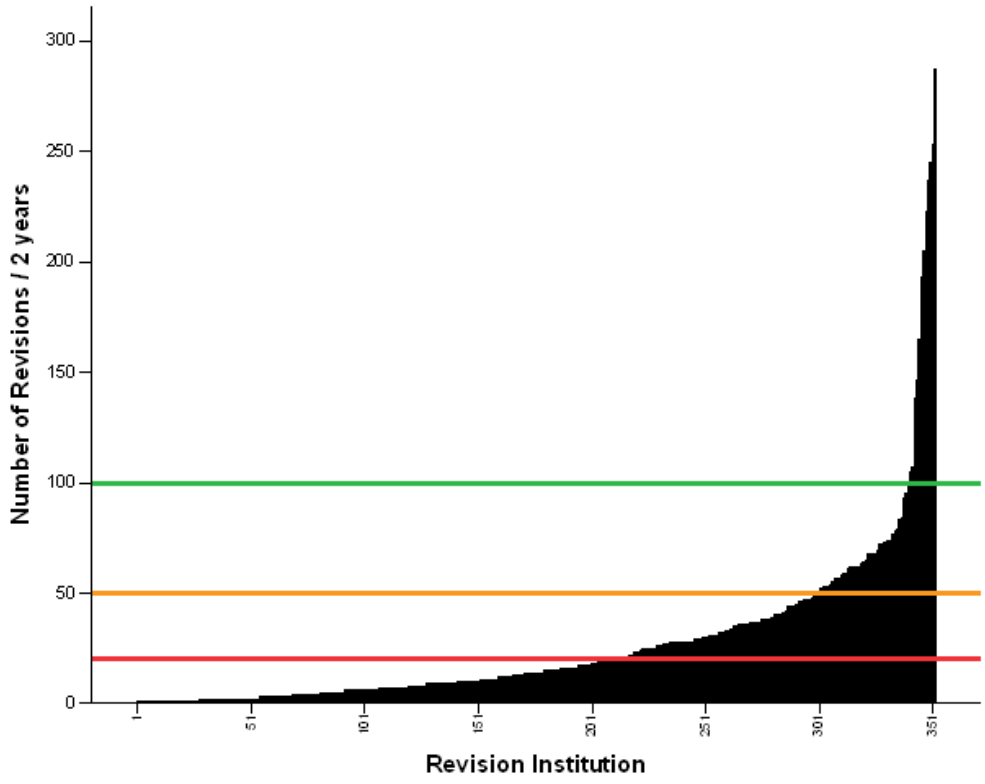


Figure 6.2: Bar chart of the number of revisions performed between 1st July 2008 and 30th June 2010 in the 359 institutions entering revision knee data on to the NJR, ordered by revision volume. Line key: Red = 20 revisions, Orange = 50 revisions, Green = 100 revisions

Number of revisions per 2 years	Number of centres	Centre type	Percentage	Cumulative Percentage
1	28	Independent - 21 ISTC - 1 NHS - 6	7.8%	7.8%
2-5	65	Independent - 43 ISTC - 2 NHS - 20	18.1%	25.9%
6-10	61	Independent - 40 ISTC - 0 NHS - 21	17.0%	42.9%
11-19	59	Independent - 24 ISTC - 3 NHS - 31	16.4%	59.3%
20-49	91	Independent - 12 ISTC - 2 NHS – 78	25.3%	84.7%
50-99	43	Independent - 1 ISTC - 2 NHS - 40	12.0%	96.7%
>99	12	Independent - 0 ISTC - 0 NHS - 12	3.3%	100.0%

Table 6.12: Number of revisions performed at each centre between 1st July 2008 and 30th June 2010 (Key: Independent = Independent Hospital, ISTC = Independent Sector Treatment Centre, NHS = NHS Hospital).

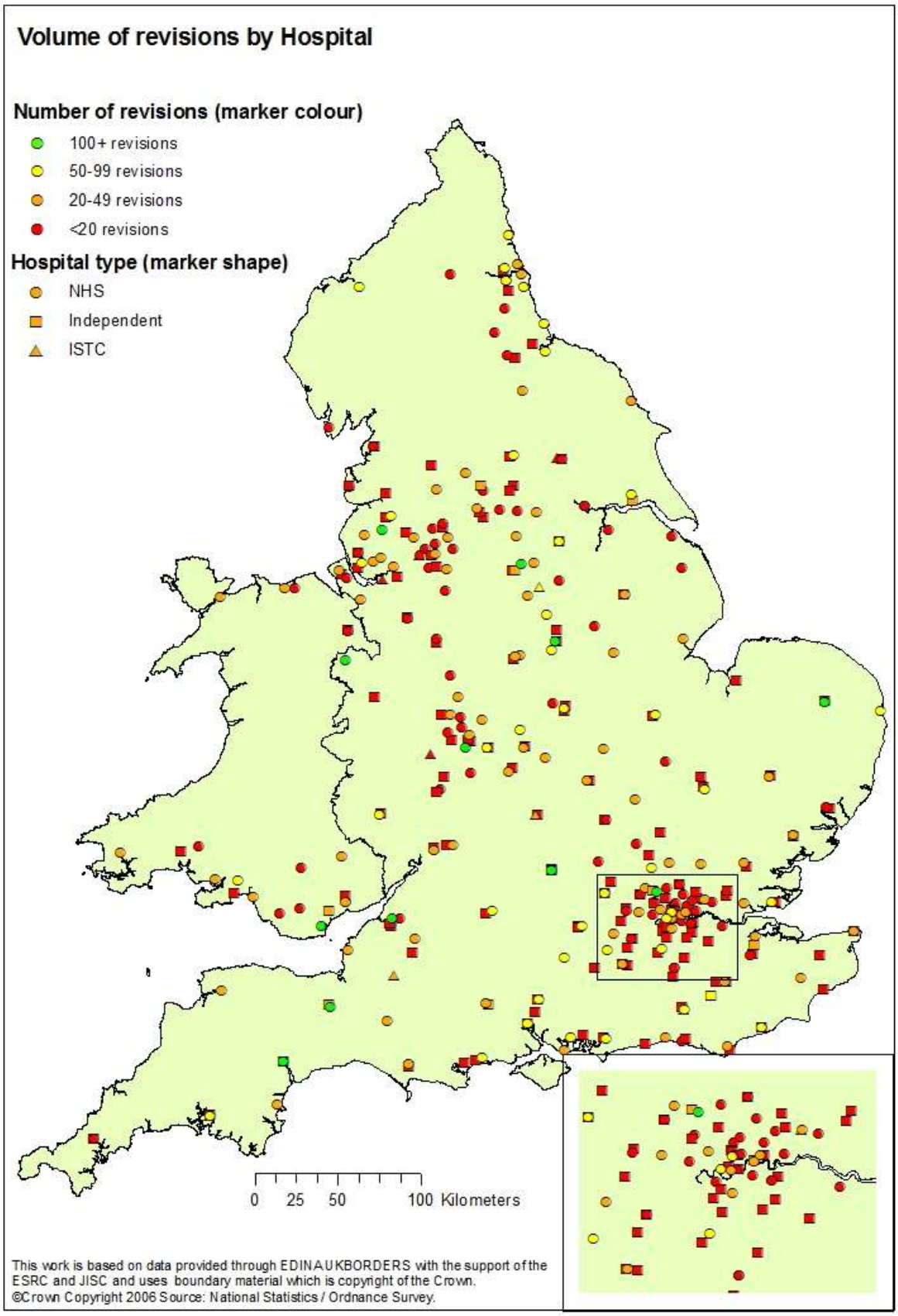


Figure 6.3: Distribution of the 359 centres performing knee revision surgery (Key given in the top left corner of the plot). Analysis based on 2 years of surgical practice.

The median number of revisions per 100 primary knee arthroplasties was 5 (Range 0 to 300). Prior to further analysis the hospital with a value of 300 was excluded as this was found to be a significant outlier representing an independent hospital that performed 2 primaries and 6 revisions during the period of analysis. There was a difference in revisions per 100 primaries dependent upon hospital type with NHS hospitals performing a greater number (Median 7, Range 0 to 76) than either independent hospitals (Median 3, Range 0 to 16) or ISTCs (Median 2, Range 0 to 6) (NHS vs. Independent and/or ISTC, $p<0.001$).

In total 80 centres were classified as "low ratio revisers" (22 NHS Hospitals, 52 Independent hospitals, 4 ISTCs). Fifty-seven centres performed >10 revisions per 100 primaries including all of the 12 highest volume hospitals possibly indicating the tertiary nature of their practice.

In total 212 centres were classified as either "low volume" or "low ratio revisers" and thus fell below the audit standards. All 80 "low ratio revisers" were also "low volumes" centres. Ninety-one per cent of all independent hospitals (128 of 141) fell below the audit standards. Of the 147 centres meeting the audit standards 130 were NHS hospitals (63% of the 208 NHS hospitals), 13 were independent hospitals (9% of the 141 independent hospitals) and 4 were ISTCs (40% of the 10 ISTCs). A final geographic plot was constructed based on the following three groups 1) Volume <20 per 2 years **and** <2.5 revisions per 100 primaries; 2) Volume <20 per 2 years **or** <2.5 revisions per 100 primaries; 3) Volume ≥20 per 2 years and ≥2.5 revisions per 100 primaries. This demonstrated that the 147 centres in the group 3 were evenly distributed within England and Wales (figure 6.4).

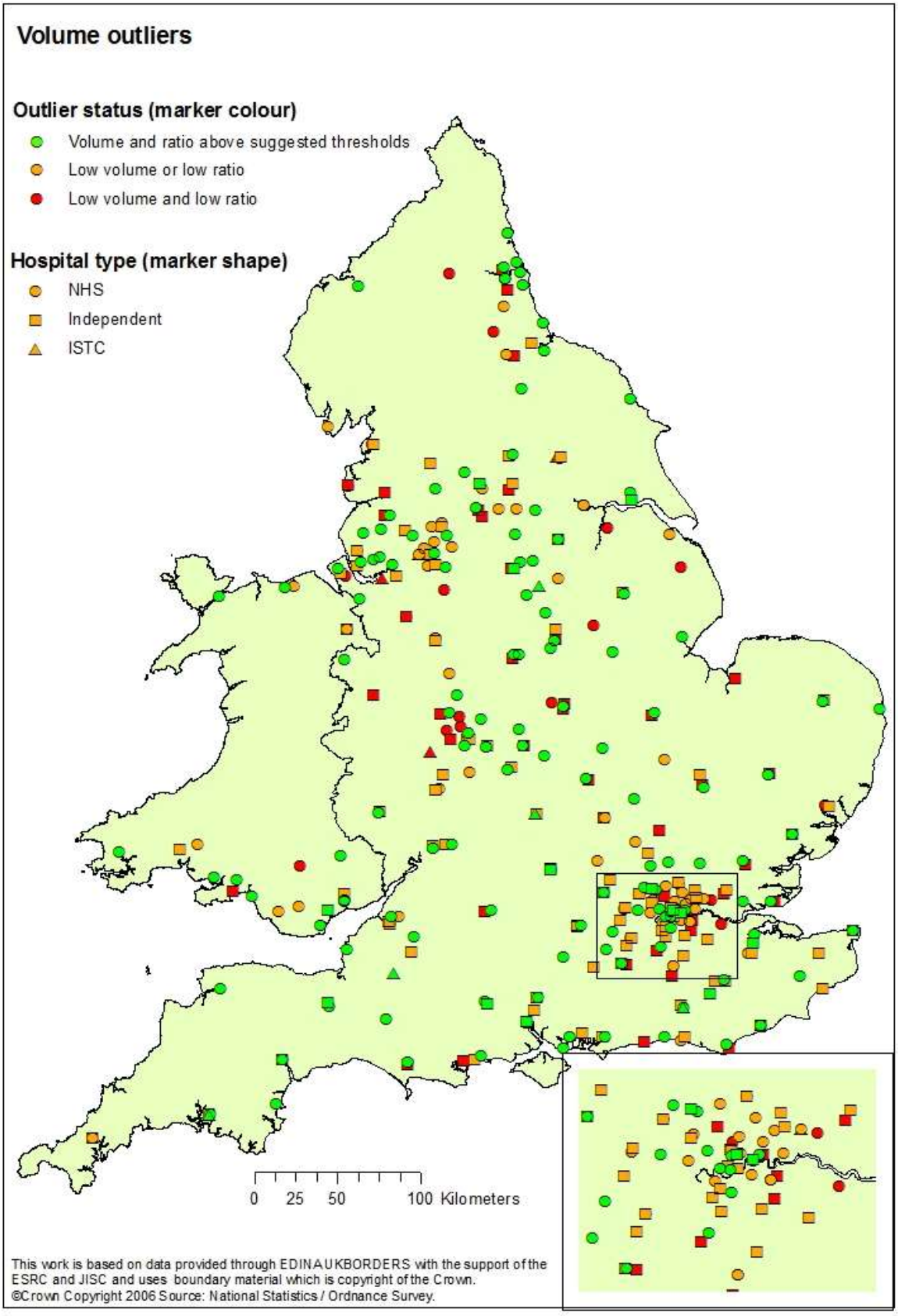


Figure 6.4: Distribution of the centres performing knee revisions dependent up their combined volume measures (see text for details) (Key given in the top left corner of the plot). Analysis based on 2 years of surgical practice.

Analysis 10: Discussion

This analysis demonstrates that a significant proportion of centres in England and Wales are performing knee revisions in small volumes. The majority of centres (213 of 359 (59.3%)) fell below one or other of our audit standards for volume (<20 revisions per 2 years or <2.5 revisions per 100 primaries) and 80 centres fell below both. This included nearly all independent hospitals. The distribution of both of the annual revision volume and the number of revisions per 100 primaries are skewed to the right, indicating that there are a high number of low volume centres and centres doing considerably more primaries than revisions.

An association between increased volume and a variety of outcomes including functional scores (Katz 2004, Katz 2007, Shervin 2007), mortality (Lavernia 1995, Judge 2006), length of stay (Lavernia 1995, Bozic 2010), complication rates (Judge 2006, Chowdry 2007), ITU requirement (Harmon 1999) and rates of home discharge (Brown 2001) has been previously reported across a range of surgical procedures. Greater familiarity and experience with complex surgical procedures increases confidence and understanding at both the individual surgeon and hospital level. In their systemic review of 163 articles, Chowdhury *et al* found that 74.2% of the papers reviewed reported significantly improved outcomes in higher volume centres (Chowdry 2007). Increasing hospital volume has been shown to improve functional results at 2 years post Total Knee Replacement (Katz 2004, Katz 2007) and reduce mortality rates after primary and revision knee arthroplasty (Lavernia 1995, Judge 2006).

This analysis poses two questions: what volume of surgery is enough? and how might services be rationalised?. A number authors have suggested values for minimum surgeon volumes (>15 to >100 TKR per year) (Hervey 2003, Katz 2004, Schulze 2006) and trust volumes (>50 TKR/THR per year) for primary arthroplasties (Judge 2006). As far as we are aware no such value for knee revision exists, largely due to methodological issues in the previous studies analysing this subset of patients (Marlow 2010). We have suggested that

hospitals should undertake a minimum of 20 revisions every 2 years (10 per year) based on the results from previously published work (Lavernia 1995, Judge 2006, Yasunaga 2009). This threshold is at the lowest end of those used in articles assessing the impact of hospital volume and therefore should only exclude the very lowest volume users. It also seems intuitively reasonable when one considers that in many centres there may be multiple surgeons each performing only 1 or 2 revisions per year. We also felt that the proportion of revisions to primaries was an important variable to consider and suggest that centres should ideally be performing >2.5 revisions per 100 primaries. We found that no centre doing >20 revisions per 2 years performed <2.5 revisions per 100 primaries which serves to vindicate this value as a cut off.

The issue of how to rationalise service poses a particular problems, especially in areas of increased population density and also in areas with a high proportion of patients who want their operation done in an independent hospital environment. One way to reduce the number of low volume centres (213) would be for hospitals within the same trust to centralise service at a local level. This would reduce the number of "low volume" centres by 38 and require only 215 revisions to be relocated. The majority of ISTCs (134 of 151 (89%)) fell below the suggested audit standards. These hospitals undertook only 1511 revisions, representing 16% of all revisions. By rationalising the use of independent hospitals and ISTCs the number of centres performing revisions would be halved. Many larger NHS facilities have access to private facilities that could accommodate the private workload whilst maintaining higher centre volumes.

Unfortunately this analysis was unable to examine the number of surgeons performing revision knee replacements and their associated distribution of surgeon level volume data. Some might argue that much of the time it is experienced high volume surgeons operating in low volume centres. The assumption being that high volume surgeons consistently achieve excellent results despite hospital volume. However, evidence suggests that high volume surgeons operating in a low volume centres are unable to attain comparable result

to those achieved by operating in high volume centres (Harmon 1999). Conversely if low volume surgeons operate in high volume centres they achieve better results than if they operated in low volume centres, albeit never as good as the high volume surgeon. This indicates that both surgeon and centre volume are interrelated (Harmon 1999). The ideal situation being a high volume surgeon operating in a high volume centre, with services rationalised to produce centres doing above a minimum threshold of procedures each year.

Analysis 10: Summary

This analysis found that a significant number of institutions are performing small volumes of knee revision procedures. To ensure safe and sustainable practice with better outcomes for patients, consideration should be given to whether this service would be better concentrated in fewer centres performing higher revision volumes.

Analysis 10: Limitations and reflection

In addition to the general methodological limitations of the NJR and PROMs datasets discussed later (section 8.2) this analysis contained a number of specific limitations.

Due to the nature of the data this analysis was entirely descriptive. Due to the small numbers of cases done in each unit and the limited outcome measures available we were unable to directly examine whether there was link between volume and outcome for revision procedures. Revision PROMs records for this cohort totalled <800 as analysis was based on procedures performed between 1st July 2008 and 30th June 2010. This equates to just over 2 PROMs records for each of the 359 centres within this analysis. Re-revision is a rare outcome, takes time to occur and is difficult to link to the previous revision procedure. These factors made it impossible to adequately examine the impact of volume upon

functional recovery, patient satisfaction and re-revision and highlight the difficultly in performing any meaningful analysis on this very heterogeneous group of patients.

Some of the issues surrounding the assessment and definition of surgical volume have previously been discussed (analysis 4, section 4.5). These are further highlighted in this analysis were, even by including 2 years' worth of data, a number of centres were only performing 1 procedure. We also had no information on surgeon volume which in our prior analysis of UKR volume (analysis 4) was shown to be a more important predictor of outcome than centre volume. It may be that the procedures done in the lowest volume centres were done by 'high volume' surgeons and therefore had the potential to achieve the best results. Due to inadequacies in the data it was not possible to examine this in any detail and this question remains unanswered.

While it is logical to assume these complex procedures do better in the hands of specialist 'high volume' centres and surgeons, the literature supporting this contention is limited and we have been unable to add significantly to it. This analysis demonstrates that there is variation in the volume of procedures performed at centres across England and Wales but what this means, how it relates to patient outcomes and how it should impact upon service delivery is a matter for debate.

Chapter 7: Overview of the main findings from the analyses performed

The key findings from each of the analyses performed are stated below:

Analysis 1: Analysis of functional improvements and rates of complications for UKR and TKR (pages 82 to 92)

- There were no demonstrable differences in the improvements in either knee-specific PROMs, general health PROMs or rates of complications between UKR and TKR.

Analysis 2: Analysis of the reasons for revision and rates of failure for unexplained pain for UKR and TKR (pages 93 to 101)

- The overall rates of revision and rates of revision for unexplained pain are significantly higher for UKR when compared to TKR.

Analysis 3: Analysis of the reasons for revision and rates of failure for medial and lateral UKR (pages 102 to 117)

- The rates of survival and reasons for revision for medial and lateral UKR are similar.

Analysis 4: Analysis of the effect of centre and surgeon operative volume on UKR rates of failure (pages 118 to 134)

- UKR failure rates are lowest when surgery is performed by specialist, high volume centres and surgeons.

Analysis 5: Analysis of the influence of surgical factors on PROMs (pages 141 to 153)

- Of the surgical factors analysed only implant brand and hospital type were seen to influence the magnitude of the PROMs improvements. However the effects of these

factors were small when compared to patient factors, and in particular the pre-operative PROMs and general health status.

Analysis 6: Analysis of the relationship between pre- and post-operative factors and patient satisfaction with TKR (pages 154 to 167)

- The most important determinants of satisfaction are the patient's perception of the success of their operation and post-operative function. Pre-operative variables have a minimal influence upon post-operative satisfaction.

Analysis 7: Analysis of the relationship between Body Mass Index and PROMs (pages 168 to 178)

- Although increasing BMI is associated with poorer pre and post-operative PROMs, the improvement experienced by patients is similar irrespective of BMI.

Analysis 8: Analysis of the reason for revision upon early PROMs following revision knee replacement (pages 184 to 197)

- Revision knee replacements only experience a fraction of the functional improvements experienced after primary total knee replacement. The magnitude of the PROMs improvements and level of patient satisfaction are dependent upon the reason for revision with revision for unexplained pain and stiffness having the worst functional outcomes.

Analysis 9: Analysis of functional outcomes after one and two stage revision for the infected knee replacement (pages 198 to 207).

- The functional outcome after revision for infection is poor. There is no difference in functional outcome between one and two stage revisions surgery in this setting.

Analysis 10: Analysis of centre operative volumes for revision knee replacements performed in England and Wales (pages 208 to 218)

- A significant number of institutions are performing small volumes of knee revision procedures

Chapter 8: Discussion

The analyses performed in chapters 4, 5 and 6 demonstrate the versatility of the combined NJR and PROMs data to answer a range of pertinent clinical questions. These large national datasets contain a wealth of data. Research using the information held within them, while observational in nature, has many strengths. There are, however, a number of general limitations to the analyses performed, a number of which have already been discussed. These include specific deficiencies in the data held within these datasets, which is a problem that is not isolated to registry data. In addition there are issues relating to the methods of data collection, as well as the previously discussed problems inherent when performing research using observational data (see section 1.4). The specific strengths and weakness of the NJR and PROMs datasets are discussed in the following sections. Particular emphasis is given to their deficiencies and merits relative to more traditional pragmatic clinical trial designs.

8.1 Strengths of the NJR/NJR-PROMs dataset

- **Size of the datasets available for analysis**

The analyses presented involved cohorts that are an order of magnitude greater than anything that could be achieved in a prospective randomised clinical trial. The size of both the NJR dataset (>400,000) cases and the NJR-PROMs dataset (>25,000 cases) made it possible to use a variety of different statistical approaches and perform analyses of rare events with high statistical power. Using national cohorts of this size moves interpretation away from sampling and towards population or census reporting, a magnitude of sampling where statistical tests of goodness of fit are not as relevant as the truth is in effect known (Burnham 2002).

The size of the NJR dataset allowed us to investigate the influence of a number of different factors upon implant revision. Analysis 4, for example, demonstrated an important association between surgeon and centre operative volume. It is unlikely the relationship observed in this analysis could have been appreciated without using a dataset of this size. Furthermore it is impossible to imagine how the impact of operative volume upon outcome could have been investigated using a prospective clinical trial design, or indeed in any other way. A dataset of this size is needed as both the explanatory variable in question (operative volume), and the event in question (revision) necessitate large numbers to allow the relationship to be appreciated.

The size of the NJR dataset also allowed us to gain an insight into infrequently performed procedures, such as lateral UKRs and revisions. In analysis 3 the revision rates, the factors influencing the revision rates and the reasons for revision for laterally implanted UKR were investigated and compared to those seen with medially implanted UKR. UKR make up approximately 9% of all knee replacements performed annually in England and Wales (NJR-AR 2012) and this analysis found that only 6% of these are implanted in the lateral compartment. Lateral UKR therefore represent only 0.5% of the 80,000 knee replacements

performed each year. As this procedure is performed relatively infrequently any RCT involving medial and lateral UKR would require the co-operation of a number of different centres and surgeons to achieve the number of lateral UKR needed for a meaningful comparison to be made. This would make it logistically complex, expensive and time-consuming. A prospective trial would, in theory, have the ability to incorporate the collection of other pertinent outcomes (e.g. functional outcomes) in its design. However, as the PROMs project matures and the number of records for lateral UKR increases, these data will become available. Functional data for these groups could be extracted directly from an NJR-PROMs dataset using the methods used in analysis 3 to determine implant laterality, and used in analysis 1 to compare functional outcomes for two different implant types. Unfortunately the small number of PROMs records presently available for the lateral UKR group (n=21) meant comparison of these groups was not possible within the analyses presented here. The additional clinical information that a RCT would provide, given the major drawbacks of this type of research design to address this clinical question, make it difficult to imagine a role for the RCT in this setting.

Revision is a rare outcome as most implants fail at a rate of less than 1% per year. For a prospective clinical trial to detect a meaningful difference in an event of this type requires the recruitment of large number of patients. While difficult, this isn't an insurmountable hurdle to overcome. When, however, interest lies in subgroups that are undergoing revision for differing reasons the event of interest becomes even rarer and the numbers needed increase substantially. These problems are circumvented by the size of the NJR dataset. Analysis 2 demonstrated the differing modes of failure observed with TKR and UKR and allowed comparative survival analyses to be performed dependent upon the reasons for revision. Similarly analysis 3 examined the differing reason for revision between medial and lateral UKR. It is only with an observational dataset as large as the NJRs that such analysis could be performed.

The size of the NJR-PROMs dataset, while smaller than the NJR dataset, also conferred a number of benefits. The PROMs project is still in its infancy, having commenced in April 2008, and has up until now been designed to be a sample analysis of both primary and revision knee replacement procedures. Due to the number of primary and revision knee replacement procedures undertaken annually in England and Wales there are, however, still a substantial number of PROMs suitable for analysis demonstrating the power of PROMs collection linked to a national registry. The number of patients available for analysis has allowed us to not only answer specific research questions but also to go further and study specific subgroups and identify important predictors of response to a number of different interventions. This is demonstrated in analyses 1, 5 and 6 where statistical modeling techniques were used to determine how a variety of covariates influence functional improvements and patient satisfaction after knee replacement. The size of the dataset allowed us to quantify the size effects of these covariates and better understand their relative influence upon the outcomes of interest. Having quantified these effects we were then able to adjust the analyses to account for any potential confounding factors. It allowed information on a large number of decisions, interventions and outcomes to be assessed simultaneously. The pragmatic design of the RCT does not readily allow such associations to be identified (Sox 2009).

In similarity to the NJR dataset, the NJR-PROMs dataset is also sufficiently large enough to allow the outcome of infrequently performed procedures to be analysed. This is demonstrated in analyses 8 and 9 in which the functional outcomes of revision procedures performed for a variety of different reasons were compared. Over the next 5 years both the NJR and PROMs projects will further mature. The NJR led 'extended PROMs' project will provide additional functional data for knee replacements through the collection of PROMs at additional post-operative time points. This will provide a substantial volume of information that could be used to assess a larger number of clinical outcomes, giving us the ability to examine in even greater detail the effects of different approaches to surgery.

- **National sampling**

National observational data provides information about the performance of different approaches to surgery in the 'real-world'. Data recorded within the NJR reports contemporary national practice and reflects what is happening in the generality of the National Health Service. The NJR, by its very nature, tries to capture data on all patients. It does not have any exclusion criteria and as such it includes patients and surgeons often omitted from clinical trials.

In contrast clinical trials often have strict inclusion and exclusion criteria which can create a false environment that is not representative of everyday practice. There is therefore uncertainty as to whether the benefits achived by average patients in these trials can be extrapolated to the average patient undergoing routine clinical care (Rawlins 2008). Under or over-representation of certain groups within the trial framework may artificially under or over estimate the size effect of the outcome under investigation. This is demonstrated in analysis 1,5 and 6 which demonstrated the significant impact a variety of patient related factors have upon outcome. Randomisation helps to eliminate bias between comparative groups within RCTs. It ensures the patient cohorts under inverstigation are similar, but it does not ensure these cohorts, or indeed the surgeons performing the interventions, are representative of the national popluation. Results from surgical enthusiasts are not always representative of the results attainable by the average surgeon (Labek 2010), but it is the results of the average surgeon that are of the greatest interest. The generalisability or external validity of clinical trials is largely dependent upon their inclusion and exclsuion criteria. How representative and applicable to national practice their conclusions are is a matter of judgement (Altman 2001).

A further advantage of national sampling is the ability to 'map' trends in current practice. Analysis 10 demonstrates how this can be achieved by geographically mapping the volume of revision TKR prcedures perfromed in England and Wales. Analyses of this type provide

important imformation about the provision and organisations of service which are of direct interest to helthcare commisioning bodies. This information can then be linked with other epidemiological population data such a social demographics, deprivation scores and population densities to examine how these procedures are utilised within different social and geographical settings.

- **Standardisation of PROMs follow up**

The PROMs data used for these analyses has the advantage of using standardised prospective collection methods with the inclusion of both pre- and postoperative data. Simple cross-sectional collection of PROMs from existing registries is problematic as patients will be at different periods in there follow-up dependent up the date when their procedure was performed. While the timing of the post-operative PROMs sampling is contentious (Kay 2011) the way in which the project has been designed ensures that the majority of post-operative questionnaires are collected at equivalent time-points. By employing inclusion criteria linked to the time to follow up and limiting the minimum and maximum periods of follow up the confounding effects of this variable can be negated. As is demonstrated in a number of our analyses even when the minimum follow-up is limited to 6 months the number of PROMs records available is still large enough to permit these analyses to be performed.

- **Standardisation of PROMs metrics**

The PROMs employ the OKS and the Euroqol-5D for the assessment of knee specific function and generic health/health utility respectively. Each of these measures was chosen based on a review of their reliability, validity and responsiveness (Browne 2007). These measures, alongside questions relating to complications, operative success and patient satisfaction, form the standard PROMs assessment for all knee replacements procedures. Previous studies investigating knee replacement have used a variety of different outcome tools (Dunbar 2004). This makes direct comparison and amalgamation of results for the

purpose of meta-analysis difficult. Standardised PROMs data collection allows the results of different approaches to surgery, different unit performance and different implant performance to be directly compared. In addition the inclusion of the generic health measures within other PROMs questionnaires (hip replacement, hernia repair, varicose vein surgery) also allows the value of procedures for different surgical problems to be directly compared.

Furthermore, whereas other outcome studies have used surgeon derived outcome measures, the PROMs projects focusses exclusively on patient reported outcomes. By using patient reported outcomes, PROMs concentrate on outcomes of importance to patients and eliminates the bias in reporting that can often accompanying surgeon derived outcomes. This is in line with current NHS reform (Darzi 2008) and mirrors the approach to data collection advocated by a number of national funding bodies (ARUK 2012).

- **Time and cost**

Prospective clinical trials are time consuming. They take time to develop and design and in many cases require funding prior to commencement. It can therefore take years for a project to get started. Currently nationally funded RCTs examining knee replacement (KAT and TOPKAT trials) are scheduled to take 15 and 10 years for recruitment and publication respectively. The TOPKAT trial has been undertaken to better define the role of the UKR, a debate which rages today. However, we will have to wait until 2020 for the results of this work to become available. These two RCTs are being undertaken at a combined cost of >£4 million.

In contrast, the time and costs needed to undertake analyses using registry data are smaller. The analyses presented here (and a number of other analyses) were all produced within a 24 month period. This includes the 6 months it took to set up the projects, put an

appropriate analytical team in place, and apply for and access the data. The costs of these analyses are difficult to quantify. NJR data entry is now mandatory with each institution responsible for the costs associated with collection and input of their own data. The collection of PROMs data is a national initiative funded through the Department of Health. These costs, along with the costs related to the management and administration of these datasets, are hidden from the researcher using registry data. The only direct costs to the research team for these analyses were the £1081 payable to the NHS information centre as a one off fee for extraction of the PROMs data and the salary of one of the research team members.

- **Standardised statistical methods**

The recent publication of guidance on the statistical reporting of registry outcomes is a step forward in the attempt to standardise reporting from registries (Ranstam 2011). Standardised methodology ensures results from different registries and healthcare systems are as comparable as is possible given the inherent demographic differences in the population that they serve. The guidelines provide a framework for on-going and subsequent registry projects which we have attempted to adhere to wherever possible.

8.2 Limitations of the NJR/NJR-PROMs dataset

- **Logistics of data handling**

Access to the PROMs and NJR datasets requires approvals from both the NHS information centre and the NJR research committee. Unfortunately there is significant disparity between the request processes and requirements for these two organisations which presents an obstacle to data access. Each process has their own requirements for data access and storage which include verifying the site and security arrangements for the required data, provision of a named Caldicott guardian, completion of information governance and data handling/re-use agreements. Although there is significant overlap in the processes for each application the two are quite separate meaning repetition and duplication of a number of the necessary steps. From initial application to receipt of the data can take between 6 and 12 months. This problem is compounded if projects require the use of data from other national datasets (e.g. Hopsital Episode Statistics, Office of National Statistics, Health Protection Agency etc.). This must be factored into the proposed timeframe of and potential projects. While attempts have been made to streamline the NJR side of the application process through the appointment of a research co-ordinator and a 3 week deadline for project review by the NJR research sub-committee the process for PROMs application remains unchanged.

Because the two applications are completely separate the requested data comes as two separate databases (an NJR and a PROMS dataset) that have to be linked, verified, cleaned and recoded prior to analysis. While linkage is performed by Northgate, the NJR data handler, the other elements have to be performed by the research team. This can be a time consuming and complex process for the unwary researcher. Thankfully the NJR and PROMs datasets contain a number of unique identifiers which make linkage relatively straightforward. If additional datasets are to be used then the complexity of linkage increases as key identifiers may not be present in all datasets and algorithms based on a number of different patient identifiers may need to be employed. The use of patient identifiers also adds the requirement for ethical approval, increasing the time required to

produce a workable dataset still further. An appropriate amount of time must be set aside for this process. While one of the strengths of observational registry data is that projects can be completed in a shorter time frame it must be realised that there is still a significant time requirement to ensure project completion, most of which relates to data application and preparation.

- **Issues with the NJR dataset**

Data quality:

The NJR uses three key indicators to determine its performance: compliance; consent; and linkability (see section 1.3.1). Ideally the rates for compliance, consent and linkability should be greater than 95% to allow confidence in the data that are produced. Generally, missing data of less than 5% of the total is unlikely to distort the overall results (NJR-AR 2012). Currently only the percentage of linkable procedure (95.5% in 2011/12) exceeds this benchmark, although the rates for compliance (90.3% in 2011/12) and consent (90.4% in 2011/12) are improving year on year (NJR-AR 2012).

Compliance is expected to improve still further with the introduction of recent legislation to make submission of NJR data mandatory. Amendments made in April 2011 to the Standard NHS Contract for Acute Services require all providers to participate in audits relevant to the service they provide within the National Clinical Audit and Patient Outcomes Programme (NCAPOP), of which the NJR is part. The submission of complete data to the NJR is, therefore, now mandatory for all NHS Trusts and foundation Trusts within England. The Welsh Government has also agreed that the NJR is mandatory for all NHS Wales hospitals.

While these key indicators continue to improve there is, as yet, no information about the accuracy of the data recorded in the registry. If a procedure is performed on a specific

patient on a given day by a particular surgeon it is currently unclear whether that information is accurately reflected within the records held within the registry. Other registries have audited their data recording and found concordance rates to be above 95% (Swedish registry AR 2009). However, as yet no audit of NJR data has been completed and therefore lingering concerns remain. This is an important issue to resolve if registries are to be used as tools to shape surgical decision making and as reservoirs for research.

Duration of follow-up:

The NJR started recording data in 2003 and the censor date for the NJR data used in these analyses was 31st December 2010. Accordingly the maximum follow-up for these analyses was 8 years. Because of the numbers of implants at risk at eight years was small, we were only able to assess mid-term survival and in most cases chose to focus on results at 5 years where the number of patients at risk was large and reliable estimates of survival could be made. It is possible that with longer follow-up the implants under scrutiny might exhibit differing patterns of failure and as such the analyses performed may produce alternative results. While we can therefore be confident of the observations made at 5 years we cannot make any inferences about longer term behaviour and implications for clinical practice. Currently a number of other registries are able to report 10 (AJR-AR, NZJR-AR) and 15 year survival rates (Swedish arthroplasty register), however, they do either not contain the volume of cases to allow for the analysis of smaller sub groups of interest or do not record patient reported / functional outcomes. This problem will be corrected as our own registry matures over the next 5 to 10 years. It will also coincide with a likely increase in the coverage of PROMs collection and introduction of initiatives such as extended PROMs which will provide additional functional comparison to be made, further increasing the power of the NJR.

Identification of revisions:

No method for the identification of revisions is fool proof. Identifying revisions through a registry may miss some revisions, usually because the revision is not recorded by the surgeon or operating unit or the procedure is miss-coded as a re-operation. As a consequence the true revision rate is underestimated.

These analyses used NJR–NJR linkage to establish the rate of revisions rather than NJR–HES linkage, on which previous NJR survival analyses have been based. This is consistent with the methods now used by the NJR for its annual reporting and outlier analyses. The Hospital Episodes Statistics (HES) is a national statistical database containing information on all admissions to NHS hospitals in England. It was previously used as an alternative source for the identification of revision procedures. Comparison of the merits of NJR–NJR and NJR–HES linkage suggest that whereas NJR–NJR linkage underestimates revisions to some undetermined extent, NJR–HES data are likely to overestimate revisions because of the inclusion of some reoperations which are erroneously recorded as revisions (NJR-AR 2012). It is likely, then, that the real revision rate lies somewhere between the two. This problem is only really an issue when comparing revision rates from different registries using different linkages and means of identifying revisions, or when basing clinical decisions on isolated values. It is, however, unlikely to substantially bias comparisons between groups, such as the medial and lateral UKRs analysed in analysis 3, as revisions for all groups are likely to be underestimated and hence biased to the same extent. Comparisons of the rates of failure between groups within the same registry should therefore be valid as long as a consistent method of identifying revisions is employed.

Adjustment of revision data:

To ensure data capture is as complete as possible most registries collect a refined set of information on a number of key variables (Robertsson 2007). These typically include demographic details (age, gender and a rating of co-morbidity (typically the ASA grade)),

information relating to the indication for surgery, and a number of pertinent intra-operative and post-operative variables. Patient characteristics are often used as the basis for clinical decision making and may also influence clinical outcomes (Sox 2009). This can lead to uncertainty about whether differences in clinical outcomes are related to differences in patient characteristics or the interventions being compared.

To overcome this problem most observational analyses attempt to account for differences in patient characteristics by employing statistical methods that make adjustments for differences in these factors. As the number of variables collected by registries is limited, so the number of variables that can be used for adjustment is also limited. Variables that could potentially influence the revision rate may therefore not be appreciated. This will always remain a concern with comparisons made using observational data (Sox 2009). This problem can only be overcome by randomly assigning patients to the different intervention under investigation. This simple action eliminates much of the uncertainty that envelops the interpretation of observational research (Sox 2009) and highlights the value of the RCT in defining a causal relationship between intervention and outcome.

- **Issues with the NJR-PROMs dataset**

Missing data:

As with many questionnaire based projects the PROMs project suffers with problems of missing or non-returned questionnaires. In addition, once the NJR and PROMs datasets were linked it became evident that there were a number of cases for which the date of operation stated on the NJR database was significantly different to those recorded by patients on their post-operative questionnaire. Our analysis of the linked NJR-PROMs dataset for knees suggests that the number of records lost for these two reasons may be upwards of 30%. Increasing the completion rates and accuracy of the date of operation recording should be a priority as the PROMs projects moves forward.

Specific PROMS issues:

There are a number of issues specific to the use of PROMs which transcend both observational research and randomised controlled trials. It is worth briefly discussing these so it can be appreciated that these measures have their own inherent flaws.

Firstly, interpretation of PROMs must be undertaken in the face of changing patient's expectations with the outcomes of hip and knee replacement. Care must therefore be taken when analysing variables such as patient dissatisfaction as this may be a manifestation of unrealistic expectation rather than poor outcome. It is questionable whether the information collected within the PROMs adequately captures the important issue of patient expectation (Danielson 2010). Secondly, while collapsing the EQ5D index into a single score is useful it also means losing information about the domains in which any differences in health-related quality of life occur. Thirdly, PROMs are based on the assumption that any change in health is directly linked to the operation under scrutiny without consideration of other extraneous factors that may concurrently affect the health and well-being of the patient. Fourthly, absence of improvement does not necessarily indicate an unsuccessful outcome when treating a degenerative condition. In some cases the goal of surgery may be to slow the rate of disease progression and degradation in quality of life, or to avoid a future health problem. These will not be adequately captured using PROMs. Finally, PROMs is entirely subjective which means it is not a "hard" definable endpoint. The implication of this is that it may neither be adequate nor appropriate as a basis for making important decisions about health care provision. As it is subjective it also excludes certain patient groups (dementia, severe cognitive impairment) from analysis.

The use of PROMS on a population scale is also not without problems. National sampling presents logistic issues, interpretation requires care and the use of PROMs that have been designed to sample specific populations of patients as a basis for clinical decision making is questionable (Carr 2012). Some authors have questioned whether the routine collection of

PROMs data in clinical cohorts should be used to assess the impact of health care interventions on patient outcomes and to guide resource allocation. These same authors believe that to answer such questions requires data (and more than just PROMs) from well controlled comparative studies rather than data from clinical cohorts (Dawson 2010). While the recording of national PROMs data represents a step forward the comments of these authors suggest that we may be extending the remit of these measures for purposes that they were not intended for. There is a risk that with such a valuable resource at our disposal we become over-reliant on it for all the answers.

Confounding:

In similarity to the problems seen when making comparisons using NJR data, comparisons made using PROMs data are also prone to the influence of confounding. As the PROMs project records a greater breadth of information relating to patient characteristics more detailed adjustment can be made. Using statistical modeling it was possible to account for many of the potential confounding factors. However, if data is not recorded it cannot be used within these models. Presently, variables such as mental health rating and educational status which are known to influence patient reported outcomes are not recorded and therefore cannot be used. Our regression models only accounted for approximately a quarter of the observed variation in the OKS improvement and a half of the observed variation in the EQ5D improvement (analyses 1 and 5). Inclusion of additional variables as described above may have improved these percentages and produced better explanatory models.

Validity of some outcomes in PROMs:

The key PROMs outcomes (OKS and EQ5D) are valid, responsive and consistent assessment tools (Browne 2007). However a number of the other assessment modalities are not. The ordinal scales used to assess satisfaction and success have not been validated but mirror similar adjectival scales used for assessing patient reported satisfaction in national cohorts

(Robertsson 2001, Baker 2007, Bourne 2010). The benefit of these scales is that they give a simple representation of the patient's perceptions of the results of surgery. It is also unclear how useful patient reported complications rates are, particularly if there has not been any attempt to verify this information against hospital records.

PROMS follow-up:

Analyses based on data from established national databases such as the NJR and PROMs are constrained by their design. This includes the type, quantity and timing of data collected. When setting up a prospective clinical trial these aspects of data collection are tailored to the research question. The question comes first and the data comes second. In observational registry research the opposite is true, the data comes first and the question comes second.

This being the case, our PROMs analyses had to be based on solitary 6 month follow-up data as this is what the PROMs project currently collects. Inferring benefit based on two PROMs observations (one pre-operative and one post-operative) makes the timing of the second observation crucial. Whereas the majority of functional improvement after knee replacement occurs within six months, further small improvements can be expected up to a year post-operatively (Kay 2011). These analyses therefore do not consider the differences that may appear with continued follow-up.

The NJR is currently performing a programme of extended PROMS sampling in a 'one off' cohort of approximately 10,000 patients. This will involve regular PROMs assessment out to 5 years post-operatively. However, due to the financial and logistical contraints of national sampling there are currently no plans to make this part of the rountine PROMs collection. The analyses presented have attempted to limit the effect of the short follow-up by ensuring a minimum follow-up of 6 months. By this time point the majority of improvement

in functional scores has already occurred and scores are starting to plateau (Judge 2011). In addition these analyses were not intended to report the final improvements gained following surgery, but to compare outcomes between different cohorts at an equivalent timepoint. The presence of a difference in the improvement in scores at 6 months is therefore valid, especially as the scores are likely to improve only minimally thereafter.

Representativeness of PROMs:

To produce a useable NJR-PROMs dataset a number of pertinent exclusions were employed. The included data was then compared to both the excluded data and data for the total population of patient undergoing knee replacement using the NJR annual report (NJR-AR 2012). This suggested that, other than a small difference in the ASA grades, the patient characteristics of these groups were well matched. The final TKR and UKR cohorts were also similar to those patients undergoing these two procedures according to the NJR annual report (NJR-AR 2012). The final NJR-PROMs cohort therefore seemed to be a representative sample of the wider population undergoing knee replacement procedures.

While their demographics were similar the number of UKR procedures was smaller than expected. UKR makes up 9% of all knee replacements entered on to the NJR but within the NJR-PROMS dataset is represented by 2% of primary knee procedures (analysis 1). The reason for this is not entirely clear. It may be related to the fact that the initial collection of PROMs data was restricted to NHS funded procedures. As a greater number of UKR than TKR procedures are privately funded and undertaken in independent rather than NHS hospitals these patients may not have been invited to participate in PROMs collection. While this may explain some of the discrepancy it is unlikely to explain it all. Questions about whether this sample is a true representation of all patients undergoing UKR must therefore remain.

8.3 Conclusion

National knee arthroplasty registers contain a wealth of information. They were originally designed as tools for auditing the performance of different approaches to knee surgery at a national level, but are increasingly being used as a conduit for clinical research. The NJR is now the largest registry in the world and as it matures with the collection of 10 years data it will become an increasingly powerful aid to clinical decision making. The recent addition of PROMs data, which can be linked to the NJR through a straightforward linking mechanism, represents a significant step forward in terms of the assessment of patient outcomes. It is a move away from revision rate as the predominant method by which registries assess outcome following surgery. This was a major criticism of registry analysis in the past (Wylde 2011).

The ability to link PROMs data to the NJR will, for the first time, allow researchers to look at how different patient groups view the results of the various different surgeries they are offered. The limitations of data such as this must not, however, be underestimated. Making decisions based on PROMs data is problematic and it is important that patients are not denied access to successful treatments because a small minority is seen to benefit less that the majority. While there are many advantages to registry based research, it is clear that this type of research can never replace prospective clinical trial designs. As the current chair of the NJR editorial board and president of the British Orthopaedic Association states:

'It is important to point out that registry data, whilst very important, needs to be supplemented by other prospective studies and randomised controlled trials to provide a broader picture." (M.Porter 2012)

Randomised controlled trials, despite their own inherent limitations, are still the best way of proving a causal relationship. Well-designed trials limit confounding and bias and remove

the uncertainty of surgical decision making using randomisation. While it is possible to limit these problems in observational datasets through statistical modelling, they will always remain to some degree.

Recent initiatives, such as the comparative effectiveness research movement, increasingly appreciate that it is only by using data from both observational databases and clinical trials that robust recommendations relevant to current clinical practice can be made. The key question researchers have to answer now is how these two approaches can be integrated, how we can maximise the value of our registry data and in what situations should a clinical trial be used in preference to registry analysis and vice versa?.

8.4: Reflection of work undertaken as part of this thesis

The completion of this medical doctorate has helped me develop both professionally and personally. Time spent in full time research has allowed me to build on the skills and knowledge gained during my previous Masters in Evidence Based Practice (Orthopaedics). Taught modules at Newcastle University covering critical appraisal, literature searches, database creation and management and paper writing have proved invaluable. They have supplemented the practical experience gained by working directly with, and subsequently writing and publishing work from, the NJR and PROMs datasets. Prior to starting this thesis I was already enrolled in a postgraduate diploma in statistics course. Specific modules on medial statistics and generalised linear modelling allowed me to be confident in the statistical analysis I was performing and gave me the confidence to undertake the majority of the analyses presented in this thesis independently. Based on my experience, I would suggest that some form of formal statistical training is essential for future researchers wanting to work with datasets of this size.

The initial research requests submitted to the NJR pre-dated the decision to undertake this thesis and my time as the NJR research fellow. At that time I was unaware of the magnitude of the datasets, the level of analysis required and I hadn't constructed my research team. I consequently made the mistake of only naming a small number of research collaborators within these requests and on the subsequent data re-use and data storage agreements. This meant that when I received the 'raw' data I was the only person able to access, view and analyse it. This left me somewhat isolated. It made it difficult to seek help when I had problems manipulating or interrogating the data as I was unable to show someone exactly what the problem was. Fortunately the NJR had employed two research fellows at this time and this meant I could share problems with the other fellow, Simon Jameson. As he was working with the same datasets we were often able to help one another to resolve problems and offer advice and guidance when problems were encountered. Without his support I would not have been able to complete a number of the analyses presented in this

thesis. If I were to start again I would construct my research team prior to making any data requests, ensuring they had the necessary skill mix to help with a range of database and statistical issues. In addition I would definitely ensure they were named on the NJR and PROMS data-reuse agreements so that they could properly support the research process.

One of the great things about registry data is that once you have it in your possession you are ready to start working with it. Experiments do not need be conducted, results to not need to be awaited and patients do not need to be recruited and followed up. This makes it very easy to develop and test out ideas for research projects. For example, the BMI analysis presented in analysis 7, arose from a simple e-mail from one of my clinical supervisors asking 'is there a link between BMI and functional outcome?'. Two hours later I was able to reply that 'yes, there did seem to be' and from this simple exchange analysis 7 was developed. I was able to work in this way due to my position as the NJR fellow. All research output was subject to review by the NJR research committee to ensure the data had been used appropriately; the analysis was of sufficient quality and most importantly that it did not overlap, replicate or conflict with the work of other researchers using the registries data. In situations where there was overlap analyses would have had to have been reworked or abandoned. Thankfully this was not the case for any of the work presented in this thesis. The ability to work in this way and get immediate answers was a major advantage of the NJR fellow post. It is undoubtedly the reason why I have been able to produce such a large volume of work in such a short space of time. Researchers hoping to replicate this level of activity must bear in mind that they will probably not be afforded the same degree of latitude when working with the registries data.

This work has given me the opportunity to travel and present at a range of national and international meetings. The feedback from colleagues about the work has been overwhelmingly positive and demonstrates the interest there is for registry based observational research and the value clinicians attribute to this type of work. However, the more I reflect on this work the more I realise its limitations. This has tempered the initial

enthusiasm I had when working, presenting and publishing using this data. I have only come to realise the problems and limitations of these datasets by spending over 2 years working with them. These limitations are not appreciated by the majority of clinicians. Having spoken to clinicians and answered questions about these analyses it is clear that many of them are using them as the basis for clinical decision making without fully appreciating their flaws. For this reason researchers using large observational datasets such as the NJR and PROMs need to be openly critical of their work and report responsibly so that results can be placed in the context of the data. I now wonder whether some of the bold statements made in early papers and presentations can truly be justified and consequently take a much more critical view of the results I present. While there is undoubted value in this work it does not give us all the answers. Where possible it needs to be supplemented by well designed clinical trials.

A number of possible areas for future work have already been discussed within the limitations and reflections sections of each of the individual analyses. This piece of work has shown that registries can produce meaningful research output. But how this compliments research from clinical trials remains unclear. Future work needs to compare and contrast the strengths and limitations of registry research to those encountered when using prospective randomised clinical trial designs. In this way we might be able to better understand how and when these differing approaches should be used, which will allow us to develop effective and cost-efficient research strategies for the investigation of arthroplasty outcomes.

Chapter 9: Appendices

9.1: Appendix i: NJR Research application form

APPLICATION FORM **NJR Ref:**

Application to Use Data from the National Joint Registry (NJR) for the Purposes of Research

Study Title:

(1) Chief Investigators

(2) Statistician responsible (Names):

(3) Has ethical approval been sought for this application?
IF YES, approved or not (please provide details and progress):
IF NO, reasons why not

(4) Information Governance: Please provide information on your policies including data storage, security, transfer and destruction, along with the monitoring processes:

Has approval from the National Information Governance Board (NIGB) been obtained? Please give details : (See above)

(5) If you have requested access to patient identifiable data please give your rationale:

(6) Does this study involve linking to patient identifiable data from other sources? If yes, please give details:

(7) Does this study involve requesting any additional information from health professionals or patients? If yes, please give details:

(8) Please provide details of the financial sponsor for this work:

(9) Please provide details of any conflict of interest:

(10) List a minimum of 5 Key Words:

(11)Have you applied previously? If yes, please give details:

(12)Please provide details of funding with regards to meeting the cost of this application:

(13) Research Protocol (Maximum 2 sides of A4 in total):

- Lay Summary of the study (maximum 100 words):
- Background, timeframe and rational of the research:
- Objectives and expected outcome of the research:
- Study type (hypothesis testing, generating or both):
- Study design (Including justifications):
- Study population (persons, place, time period, exclusion and inclusion criteria for data required):
- Data required:
- Data fields required:
- Selection of comparison group(s) or controls – if applicable:
- Sample Size /power calculation:
- Exposures, outcomes and covariates (please describe the strategies and data sources for determining the main exposures, key health outcomes and all other variables relevant to the study objectives):
- Data/statistical analysis (outline the methodology to be used for data management and the statistical approaches to be used in data analysis):
- Project timeline:
- Patient/user group involvement:
- Limitations of study design, data sources and analytical methods:

(14) Plans for disseminating and communicating study results:

(15) Contribution to overall knowledge base /healthcare of patients:

NJR Research Request Specification

A. Joint

- Hip ☐
- Knee ☐
- Ankle ☐
- Shoulder ☐
- Elbow ☐

B. Procedure Type

- Primary ☐
- Revision ☐
- Linked Primary – Revision ☐

C. Patient Demographics

- Gender ☐
- Age at procedure ☐
- Body Mass Index ☐
- ASA Grade ☐

D. Provider

- Strategic Health Authority ☐
- Provider Type ☐
- Trust/Company ID ☐
- Unit ID ☐
- Consultant ID ☐
- Lead Surgeon ID ☐
- Lead Surgeon Grade ☐
- First Assistant Grade ☐
- Trust Name* ☐
- Hospital Name* ☐

E. Procedure Details

- Procedure Type ☐
- Patient Procedure ☐
- Side ☐
- Indications for Primary Procedure ☐
- Reasons for Revision ☐
- Approach ☐
- Bone graft ☐
- Thromboprophylaxis ☐
- Untoward Intra-operative Events ☐

F. Implant Data

- Manufacturer* ☐
- Catalogue Number* ☐
- Detail* ☐
- Batch Number* ☐
- Brand* ☐
- Implant Type ☐

* Restricted Data

TERMS & CONDITIONS

1. The recipient agrees to be held responsible for the use of the data supplied by the NJR. Under no circumstances will the data be used for any other purpose than that specifically described in this application. There will be no release of the data in any kind to a third party without the specific written consent of the NJR.

2. The recipient accepts responsibility for data protection on data that is received from the NJR and will ensure the information is kept confidential and used for the purpose of this research application only.

3. All data received from the NJR is treated as background Intellectual property and remains solely owned by the NJR.

4. The recipient will provide a written summary of the work performed on data supplied by the NJR at regular 6 monthly intervals following receipt of the data. Each update should be received by the NJR no longer than 30 days after the period end. On completion of the project, a final written report will be sent to the NJR within 3 months of the end date. The NJR reserves the right to request an update at any stage throughout the duration of the project where it is felt there is grounds and reason to do so.

5. The recipient will provide a copy of any paper/abstract for review and comment by the NJR at least one month before submitting for publication. Any reasonable amendments requested by the NJR with regards to its data will be incorporated. The recipient agrees to cite the contribution made by the NJR in all publications arising from research on the data and will provide copies of such publications. The NJR reserves the right of veto.

6. In the event that data supplied by the NJR generates ideas, rights, processes or products of potential commercial value, the recipient agrees to enter into a separate agreement with the NJR, to be negotiated in good faith in consideration of the contribution made by each party with regards to Intellectual Property ownership and exploitation rights.

7. The recipient warrants that all ethical approval/clearance required by law to receive, house and carry out research on the data provided is in place, adhering to the Data Protection Act 1998, in accordance with the support requirement under section 251 of the NHS Act 2006 and section 136 of the Health & Social Care (Community Health & Standards) Act 2003.

8. The NJR reserves the right to terminate at any stage if there is seen to be a material breach of this agreement. The NJR will seek both compensatory and punitive damages against any unauthorised release of the data.

9. The NJR uses data collected, collated and provided by third parties. As a result of this NJR takes no responsibility for the accuracy, currency, reliability and correctness of any data used or referred to in this report, nor for the accuracy, currency, reliability and correctness of links or references to other information sources and disclaims all warranties in relation to such data, links and references to the maximum extent permitted by legislation.

   The NJR shall have no liability (including but not limited to liability by reason of negligence) for any loss, damage, cost or expense incurred or arising by reason of any person using or relying on the data within this report and whether caused by reason of any error, omission or misrepresentation in the report or otherwise. This report is not to be taken as advice. Third parties using or relying on the data in this report do so at their own risk and will be responsible for making their own assessment and should verify all relevant representations, statements and information with their own professional advisers.

   The data provided by the NJR is sourced externally. All provisions of clinical data sets are made in good faith by the NJR based on the data collected at source. The NJR retain no responsibility for the quality of the data provided at source.

10. This agreement is subject to English law.

SIGNATORIES for the Recipient

a. Statistician responsible:
   I have read the Terms & Conditions and agree to abide by them. To the best of my knowledge, all the information given on this form and my application form is accurate.

   Sign Date
   Name (printed)

b. Chief Investigator:
   I have read the Terms & Conditions and agree to abide by them. To the best of my knowledge, all the information given on this form and my application form is accurate.

   Sign Date
   Name (printed)

c. Institutional Signatory:
   I have read the Terms & Conditions and agree to abide by them. To the best of my knowledge, all the information given on this form and my application form is accurate.

   Sign Date
   Name (printed)
   Position

9.2: Appendix ii: NJR K1 and K2 FORMS

National Joint Registry www.njrcentre.org.uk	MDS VERSION 3.1 Knee Operation Form: MDSv3.1 K1 v1.3
K1 Knee Primary	Patient Addressograph
Important: Please tick relevant boxes. All component stickers should be affixed to the accompanying 'Minimum Dataset Form Component Labels Sheet'. Please ensure that all sheets are stapled together.	

All fields are Mandatory unless otherwise indicated

REMEMBER! MAKE A NOTE OF THE NJR REFERENCE NUMBER WHEN YOU ENTER THIS DATA	NJR REF:

PATIENT DETAILS				
Patient Consent Obtained	Yes ☐	No ☐	Not Recorded ☐	
Patient Hospital ID				
Body Mass Index (enter either H&W OR BMI OR tick Not Available box)	Height (IN CM) Weight (IN KG)	BMI	Not Available ☐	

PATIENT IDENTIFIERS				
Forename				
Surname				
Gender	Male ☐	Female ☐	Not Known ☐	Not Specified ☐
Date of Birth	DD/MM/YYYY			
Patient Postcode			Overseas Address ☐	
NHS Number (if available)				

OPERATION DETAILS				
Hospital				
Operation Date	DD/MM/YYYY			
Anaesthetic Types	General ☐ Regional - Epidural ☐		Regional – Nerve Block ☐ Regional – Spinal (Intrathecal) ☐	
Patient ASA Grade	1 ☐	2 ☐	3 ☐	4 ☐ 5 ☐
Operation Funding	NHS ☐	Independent ☐		

SURGEON DETAILS					
Consultant in Charge					
Operating Surgeon					
Operating Surgeon Grade	Consultant ☐	SPR/ST3-8 ☐	F1-ST2 ☐	Specialty Doctor/SAS ☐	Other ☐
First Assistant Grade	Consultant ☐		Other ☐		

KNEE PRIMARY PROCEDURE DETAILS							
Side	Left ☐	Right ☐					
Indications for Implantation (select all that apply)	Osteoarthritis	☐	Rheumatoid Arthritis	☐			
	Avascular Necrosis	☐	Previous Trauma	☐			
	Other Inflammatory Arthropathy	☐	Other	☐			
	Previous Infection	☐					
PRE OPERATIVE RANGE OF MOVEMENT							
Fixed Flexion Deformity (degrees)	Less than 10 ☐	10 to 30 ☐	Greater than 30 ☐	Not Available ☐			
Flexion (degrees)	Less than 70 ☐	70 to 90 ☐	91 to 110 ☐	Greater than 110 ☐	Not Available ☐		

SURGICAL APPROACH							
Patient Procedure	Primary Total Prosthetic Replacement Using Cement ☐ Primary Total Prosthetic Replacement Not Using Cement ☐ Unicondylar Knee Replacement ☐ Patello-Femoral Knee Replacement ☐ Primary Total Prosthetic Replacement Not Classified Elsewhere (eg Hybrid) ☐						
Consultant in Charge – Default Technique used?	Yes ☐ No ☐ If Yes, ensure the relevant Consultant Default Technique is recorded on the Data Entry system. The Consultant's Default Technique is made up of several data fields.						
Approach	Medial Parapatellar	☐	Mid-Vastus	☐			
	Lateral Parapatellar	☐	Other	☐			
	Sub-Vastus	☐					
Minimally Invasive Technique Used?	Yes ☐	No ☐					
Computer Guided Surgery Used?	Yes ☐	No ☐					
THROMBOPROPHYLAXIS REGIME (intention to treat)							
Chemical	Aspirin	☐	Warfarin	☐	None ☐		
	LMWH	☐	Direct Thrombin Inhibitor	☐			
	Pentasaccharide	☐	Other	☐			
Mechanical	Foot Pump	☐	Other	☐			
	Intermittent Calf Compression	☐	None	☐			
	TED Stockings	☐					
BONEGRAFT USED							
Femur	Yes ☐	No ☐					
Tibia	Yes ☐	No ☐					
SURGEON'S NOTES							

INTRA OPERATIVE EVENT				
Untoward Intra Operative Event	None	☐	Ligament Injury	☐
	Fracture	☐	Other	☐
	Patella Tendon Avulsion	☐		

Minimum Dataset Form - COMPONENT LABELS

1. Please affix any component labels to this sheet and ensure any extra component label sheets are attached to the main Minimum Dataset Form.
2. Ensure all component details are provided, including cement.
3. The NJR DOES NOT record the following: wire, mesh, cables, plates, screws, surgical tools, endoprostheses or bipolar heads.

National Joint Registry www.njrcentre.org.uk	MDS VERSION 3.1 Knee Operation Form: MDSv3.1 K2 v1.2
K2 **Knee Single Stage Revision** **Knee Stage 1 of 2 Stage Revision** **Knee Stage 2 of 2 Stage Revision** **Knee Conversion to Arthrodesis** **Knee Amputation**	Patient Addressograph
Important: Please tick relevant boxes. All component stickers should be affixed to the accompanying 'Minimum Dataset Form Component Labels Sheet'. Please ensure that all sheets are stapled together.	

All fields are Mandatory unless otherwise indicated

REMEMBER! MAKE A NOTE OF THE NJR REFERENCE NUMBER WHEN YOU ENTER THIS DATA	NJR REF:

PATIENT DETAILS				
Patient Consent Obtained	Yes ☐	No ☐	Not Recorded ☐	
Patient Hospital ID				
Body Mass Index (enter either H&W OR BMI OR tick Not Available box)	Height (IN CM) Weight (IN KG)	BMI	Not Available ☐	

PATIENT IDENTIFIERS				
Forename				
Surname				
Gender	Male ☐	Female ☐	Not Known ☐	Not Specified ☐
Date of Birth	DD/MM/YYYY			
Patient Postcode			Overseas Address ☐	
NHS Number (if available)				

OPERATION DETAILS					
Hospital					
Operation Date	DD/MM/YYYY				
Anaesthetic Types	General ☐ Regional - Epidural ☐		Regional – Nerve Block ☐ Regional – Spinal (Intrathecal) ☐		
Patient ASA Grade	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
Operation Funding	NHS ☐	Independent ☐			

SURGEON DETAILS					
Consultant in Charge					
Operating Surgeon					
Operating Surgeon Grade	Consultant ☐	SPR/ST3-8 ☐	F1-ST2 ☐	Specialty Doctor/SAS ☐	Other ☐
First Assistant Grade	Consultant ☐	Other ☐			

KNEE REVISION PROCEDURE DETAILS		
Procedure Type	Single Stage Revision ☐ Stage 1 of 2 Stage Revision ☐ Stage 2 of 2 Stage Revision ☐	Conversion to Arthrodesis ☐ Amputation ☐
Side	Left ☐ Right ☐	
Indications For / Findings at Time of Revision (select all that apply)	Aseptic Loosening Femur ☐ Tibia ☐ Patella ☐ Infection ☐ Dislocation / Subluxation ☐ Lysis Femur ☐ Tibia ☐	Instability ☐ Wear of Polyethylene Component ☐ Component Dissociation ☐ Pain (undiagnosed) ☐ Malalignment ☐ Peri-Prosthetic Fracture ☐ Implant Fracture ☐ Stiffness ☐ Progressive Arthritis Remaining Knee ☐ Other ☐

PRIMARY OPERATION DETAILS		
Primary Operation Date OR Year	DD/MM/YYYY Please enter Date if known	Not Available ☐
Primary Operation Hospital		Not Available ☐

COMPONENTS REMOVED (Do not complete for Stage 2 of 2 Stage Revision)		
Brand of Knee Removed		Not Available ☐

SURGICAL APPROACH (Used for Single Stage Revision & Stage 2 of 2 Stage Revision)				
Patient Procedure	Revision Using Cement ☐ Revision Not Using Cement ☐ Revision Not Classified Elsewhere (eg Hybrid) ☐			
Approach	Medial Parapatellar ☐ Lateral Parapatellar ☐ Sub-Vastus ☐ Mid-Vastus ☐		Quadriceps Turn-Down ☐ Tibial Tubercle Osteotomy ☐ Other ☐	
THROMBOPROPHYLAXIS REGIME (intention to treat)				
Chemical	Aspirin ☐ LMWH ☐ Pentasaccharide ☐		Warfarin ☐ Direct Thrombin Inhibitor ☐ Other ☐	None ☐
Mechanical	Foot Pump ☐ Intermittent Calf Compression ☐ TED Stockings ☐		Other ☐ None ☐	
BONEGRAFT USED				
Femur	Yes ☐ No ☐			
Tibia	Yes ☐ No ☐			
SURGEON'S NOTES				

INTRA OPERATIVE EVENT		
Untoward Intra Operative Event	None ☐ Fracture ☐ Patella Tendon Avulsion ☐	Ligament Injury ☐ Other ☐

Minimum Dataset Form - COMPONENT LABELS

1. Please affix any component labels to this sheet and ensure any extra component label sheets are attached to the main Minimum Dataset Form.
2. Ensure all component details are provided, including cement.
3. The NJR DOES NOT record the following: wire, mesh, cables, plates, screws, surgical tools, endoprostheses or bipolar heads.

9.3: Appendix iii: PROMs Questionnaires

NHS

Knee Surgery Questionnaire

After your operation

About six months ago you had a Knee Operation. You may remember that you agreed that we could send you an *After your operation* questionnaire. Please can you fill in this questionnaire and return it using the provided pre-paid envelope. Thank you for your help.

Q1. Is anyone helping you fill in this questionnaire?

Yes ☐ 1 No ☐ 2

If the answer is yes, please give the relationship to you of the person assisting you

Q2. What is your date of birth?

D D M M 1 9 Y Y

A question about your current home circumstances

Q3. Which statement best describes your living arrangements?

I live with partner/spouse/family/friends ☐ 1
I live alone ☐ 2
I live in a nursing home, hospital or other long-term care home ☐ 3
Other ☐ 4

Q4. Please confirm when your knee operation took place (day, month, year).

D D M M 2 0 Y Y

1

Some questions about your surgery and your health

Please mark the boxes below with a tick or numbers where appropriate. If you are unsure about how to answer a question, please give the best answer you can.

Q5. Did you experience any of the following problems after your operation? Please tick *Yes* or *No* for each problem.

	Yes	No
Allergy or reaction to drug	☐ 1	☐ 2
Urinary problems	☐ 1	☐ 2
Bleeding	☐ 1	☐ 2
Wound problems	☐ 1	☐ 2

Q6. Have you been readmitted to hospital since the operation on your knee?

Yes ☐ 1 No ☐ 2

Q7. Have you had another operation on your knee since your knee replacement surgery?

Yes ☐ 1 No ☐ 2

Q8. In general, would you say your health is:

Excellent ☐ 1 Very good ☐ 2 Good ☐ 3 Fair ☐ 4 Poor ☐ 5

Q9. How would you describe the results of your operation?

Excellent ☐ 1 Very good ☐ 2 Good ☐ 3 Fair ☐ 4 Poor ☐ 5

Q10. Overall, how are the problems now in the knee on which you had surgery, compared to before your operation?

Much better	A little better	About the same	A little worse	Much worse
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

We are interested in the knee on which you had surgery.
Tick one box for every question.

Q11. During the past 4 weeks...
How would you describe the pain you usually have from your knee?

None	Very mild	Mild	Moderate	Severe
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q12. During the past 4 weeks...
Have you had any trouble with washing and drying yourself (all over) because of your knee?

No trouble at all	Very little trouble	Moderate trouble	Extreme difficulty	Impossible to do
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q13. During the past 4 weeks...
Have you had any trouble getting in or out of a car or using public transport because of your knee? (whichever you tend to use)

No trouble at all	Very little doubt	Moderate trouble	Extreme difficulty	Impossible to do
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q14. **During the past 4 weeks...**
For how long have you been able to walk before pain from your knee becomes severe? (with or without a stick)

No pain/More than 30 minutes	16 to 30 minutes	5 to 15 minutes	Around the house *only*	Not at all – pain severe on walking
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q15. **During the past 4 weeks...**
After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your knee?

Not at all painful	Slightly painful	Moderately painful	Very painful	Unbearable
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q16. **During the past 4 weeks...**
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
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q17. **During the past 4 weeks...**
Could you kneel down and get up again afterwards?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q18. **During the past 4 weeks...**
Have you been troubled by pain from your knee in bed at night?

No nights	Only 1 or 2 nights	Some nights	Most nights	Every night
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q19. During the past 4 weeks...
How much has pain from your knee interfered with your usual work (including housework)?

Not at all	A little bit	Moderately	Greatly	Totally
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q20. During the past 4 weeks...
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
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q21. During the past 4 weeks...
Could you do the household shopping on your own?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Q22. During the past 4 weeks...
Could you walk down one flight of stairs?

Yes, easily	With little difficulty	With moderate difficulty	With extreme difficulty	No, impossible
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

By placing a tick in one box in each group (Questions 23–27) below, please indicate which statements best describe your own health state today.

Q23. Mobility

- I have no problems in walking about ☐ 1
- I have some problems in walking about ☐ 2
- I am confined to bed ☐ 3

Q24. Self-Care

- I have no problems with self-care ☐ 1
- I have some problems washing or dressing myself ☐ 2
- I am unable to wash or dress myself ☐ 3

Q25. Usual Activities
(*e.g. work, study, housework, family or leisure activities*)

- I have no problems with performing my usual activities ☐ 1
- I have some problems with performing my usual activities ☐ 2
- I am unable to perform my usual activities ☐ 3

Q26. Pain/Discomfort

- I have no pain or discomfort ☐ 1
- I have moderate pain or discomfort ☐ 2
- I have extreme pain or discomfort ☐ 3

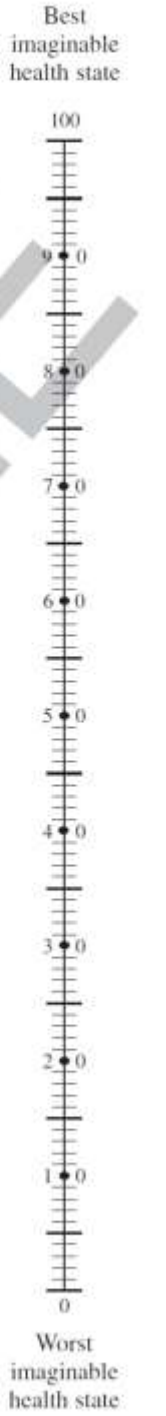
Q27. Anxiety/Depression

- I am not anxious or depressed ☐ 1
- I am moderately anxious or depressed ☐ 2
- I am extremely anxious or depressed ☐ 3

Q28. To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today.

Your own health state today



Q29. How many times have you seen a physiotherapist since you left hospital?

None	1 to 5 times	6 to 10 times	More than 10 times
☐ 1	☐ 2	☐ 3	☐ 4

Q30. Today's date (day, month, year)

D	D	M	M	2	0	Y	Y

Q31. Do you consider yourself to have a disability?

Yes	No
☐ 1	☐ 2

Thank you for your assistance.

Please return this questionnaire in the envelope provided.
You do not have to use a stamp – the postage is already paid.

Contact us for further details:

- Telephone: **XXXXXX**
- E-mail: **XXXXXXX@XXXXXX.nhs.uk**
- Write: **XXXXXXX**
- Website: **www.XXXXXX.nhs.uk**

9.4: Appendix iv: Ethics correspondence

Ethics correspondence between Paul Baker (NJR Fellow), Bill Hackett (Committee coordinator, Northern and Yorkshire research ethics committee) and Paddy Stevenson (Chairman, Sunderland research ethics committee).

From: Paul Baker
Sent: 08 September 2010 11:52
To: Bill Hackett
Subject: Chairman's advice

Dear Bill,

I am a SpR working on a number of projects allied to the National Joint Registry for England and Wales with Prof Deehan. We are confident that the proposed projects fall under the category of service evaluations and as such do not require formal ethical/NRES input. Prof Deehan, however, has suggested I contact Paddy Stevenson as a local ethics chair just to run the proposal by him to ensure that this is the case and we don't run into any trouble further down the line. Would it be acceptable to seek the Chairman's advice on the two projects I intend to undertake? If so how is the best way to contact him and what is the process for this kind of request?

Many Thanks
Paul Baker
SpR Orthopaedics
Northern Deanery

From: Bill Hackett
Sent: 08 September 2010 11:58
To: Paul Baker
Cc: Paddy Stevenson
Subject: RE: Chairman's advice

Hello Paul,

Yes that's all fine. Just email Paddy at Newcastle and I'm sure he'll respond ASAP. Paddy.stevenson@newcastle.ac.uk.

Best wishes,

Bill

From: Paddy Stevenson
To: Bill Hackett; Paul Baker
Sent: 08 September 2010 12:16
Subject: RE: Chairman's advice

Hi Paul, I would be happy to review them.

Paddy

From: Paul Baker
To: Paddy Stevenson
Sent: 14 September 2010 12:13
Subject: RE: Chairman's advice

Dear Dr Stevenson,

Many thanks for having a look at the attached proposals (Analysis of revision knee arthroplasty in England and Wales & PROMs following primary total knee replacement and unicondylar knee replacement). I think they are fairly self explanatory as they were completed on the NJR research request proforma which highlights most of the areas you will be interested in.

If you need any further information please let me know. If after reading the proposals you want to discuss them, then I can be contacted on my mobile: XXXXXXXXXXXX.

Thanks again.

Paul Baker

From: Paddy Stevenson
To: Paul Baker
Date: Tue, 14 Sep 2010 14:28:02
Subject: FW: Chairman's advice

Dear Paul

I'm happy that the first proposal 'Analysis of revision knee arthroplasty in England and Wales' does not require ethical approval / consent. With the second project 'PROMs following primary total knee replacement and unicondylar knee replacement' could you confirm no consent is being sought and all data will be collected and anonymised.

Thanks.

Paddy

From: Paul Baker
To: Paddy Stevenson
Sent: 15 September 2010 20:19
Subject: RE: Chairman's advice

Dear Paddy,

The second PROMs study will be conducted using anonymised data from the NJR which they are collecting routinely now. They obviously have their own information governance approvals etc. I have spoken to the NJR data handlers who inform me all the data can be coded/anonymised at source before it is transferred to me for analysis. I won't be seeking consent as I have no indication who the patients are or where or when they underwent surgery.

Hope this clarifies matters.

Paul

From: Paddy Stevenson
To: Paul Baker
Date: Fri, 17 Sep 2010 13:28:25
Subject: RE: Chairman's advice

Makes sense.

I agree that they don't require formal ethical review.

Paddy

Chapter 10: References and Bibliography

- Ahnfelt L, Andersson G, Herberts P. Re-operation of total hip arthroplasties. Lakartidningen 1980;77:2604-7.
- Allyson Jones C, Beaupre LA, Johnston DWC, Suarez-Almazor ME. Total Joint Arthroplasties: Current concepts of patient outcomes after surgery. Clinics in Geriatric Medicine 2005;21:527-541.
- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement: statement for reporting randomised trials: explanation and elaboration. Ann Intern Med 2001;134:663-94. American College of Rheumatology Subcommittee on Rheumatoid Arthritis Guidelines. Guidelines for the Management of Rheumatoid Arthritis 2002. Arthritis and Rheumatism 2002;46(2):328–346.
- American Society of Anesthesiologists (ASA) Physical Status Classification System. http://www.asahq.org/clinical/physicalstatus.htm
- Amin AK, Clayton RAE, Patton JT, Gaston M, Cook RE, Brenkel IJ. Total knee replacement in morbidly obese patients. Results of a prospective, matched study. J Bone Joint Surg [Br] 2006;88-B:1321-6.
- Anderson JG, Wilson RL, Tsai D, Stulberg SD, Chang RW. Functional outcome and patient satisfaction in total knee patients over the age of 75. Journal of Arthroplasty 1996;11(7):831-40.
- Antti-Poika I, Josefsson G, Konttinen Y, Lindgren L, Santavirta S, Sanzén L. Hip arthroplasty infection. Current concepts. Acta Orthop Scand 1990;61:163-169.
- Argenson JA, Parratte S, Bertani A, Flecher X, Aubaniac JM. Long-term results with a lateral unicondylar replacement. Clin Orthop Relat Res 2008;466:2686-93.
- Armitage P, Berry G. Statistical Methods in Medical Research (3rd ed) Blackwell 1994.
- Arthritis Research UK (ARUK): www.arthritisresearchuk.org/
- Ashraf T, Newman JH, Evans RL, Ackroyd CE. Lateral unicompartmental knee replacement: survivorship and clinical experience over 21 years. J Bone Joint Surg [Br] 2002;84-B:1126-30.

- Bachmeier CJ, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, Brooks PM. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. Osteoarthritis Cartilage 2001;9:137–46.
- Baker PN, van der Meulen JH, Lewsey J, Gregg PJ; National Joint Registry for England and Wales. The role of pain and function in determining patient satisfaction following total knee arthroplasty: data from the National Joint Registry for England and Wales. J Bone Joint Surg [Br] 2007;89-B:893-900.
- Bauer T, Piriou P, lhotellier L, Leclerc P, Mamoudy P, Lortat-Jacob A.Results of reimplantation for infected total knee arthroplasty: 107 cases. Rev Chir Orthop Reparatrice Appar Mot 2006;92:692-700.
- Bauwens K, Matthes G, Wich M, Gebhard F, Hanson B, Ekkernkamp A, Stengel D. Navigated total knee replacement. A meta-analysis. J Bone Joint Surg [Am] 2007;89-2:261-9.
- Bender, R. and Grouwen, U. Ordinal logistic regression in medical research. Journal of the Royal College of Physicians of London 1997;31(5):546-548.
- Bertin KC, Freeman MAR, Samuelson KM, Ratcliffe SS, Todd RC. Stemmed revision arthroplasty for aseptic loosening of total knee replacement. J Bone Joint Surg [Br] 1985;67(2):242-248.
- Biau DJ, Latouche A, Porcher R. Competing events influence estimated survival probablility: when is Kaplan-Meier analysis appropriate? Clin Orthop 2007;(462):229-33.
- Birrel F, Johnell O, Silman A. Projecting the need for hip replacement over the next three decades: influence of changing demographics and threshold for surgery. Annals of the Rheumatic Diseases 1999;58:569-72.
- Blackburn J, Qureshi A, Amirfeyz R, Bannister G. Does preoperative anxiety and depression predict satisfaction after total knee replacement? Knee 2012;19(5):522-4.
- Bohler N, Labek G. Current status of Arthroplasty Registers in Europe. European Instructional Lectures Vol 9 2009:pp 3-10.
- Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KD. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? Clin Orthop Relat Res 2010;468(1):57-63.

- Boyer P, Boutron I, Ravaud P. Scientific production and impact of national registers: the example of orthopaedic national registers. Osteoarthritis and Cartilage 2011;19:858-863.
- Bozic KJ, Maselli J, Pekow PS, Lindenauer PK, Vail TP, Auerbach AD. The influence of procedure volumes and standardization of care on quality and efficiency in total joint replacement surgery. J Bone Joint Surg [Am] 2010;92(16):2643-52.
- Brouwer RW, Raaij van TM, Bierma-Zeinstra SM, Verhagen AP, Jakma TS, Verhaar JA. Osteotomy for treating knee osteoarthritis. Cochrane Database Syst Rev 2007;18(3):CD004019.
- Brown PP, Mack MJ, Simon AW, Battaglia SL, Tarkington LG, Culler SD, Becker ER. Comparing clinical outcomes in high-volume and low-volume off-pump coronary bypass operation programs. Ann Thorac Surg 2001;72:S1009-S1015.
- Browne J, Jamieson L, Lewsey J, *et al*. London School of Hygiene & Tropical Medicine. Patient Reported Outcome Measures (PROMs) in Elective Surgery: report to the Department of Health, 2007. http://www.lshtm.ac.uk/php/hsrp/research/proms_report_12_dec_07.pdf
- Buechel FF, Femino FP, D'Alessio J. Primary exchange revision arthroplasty for infected total knee replacement: a long-term study. Am J Orthop (Belle Mead NJ) 2004;33:190-198.
- Bullens PH, Van Loon CJ, De Waal Malefijit MC, Laan RF, Veth RP. Patient satisfaction after total knee arthroplasty: a comparison between subjective and objective outcome assessments. Journal of Arthroplasty 2001;16(6):740-7.
- Burnett RS, Boone JL, Rosenzweig SD, Steger-May K, Barrack RL. Patellar resurfacing compared with nonresurfacing in total knee arthroplasty. A concise follow-up of a randomised trial. J Bone Joint Surg [Am] 2009;91-11:2562-7.
- Burnham KP, Anderson DR. Model selection and multimodel inference: A practical Information-Theoretic Approach. 2nd Edition, Springer, 2002:pp488.
- Carr AJ, Robertsson O, Graves S, Price AJ, Arden NK, Judge A, Beard DJ. Knee replacement. Lancet 2012;1-10.
- Chang CB, Yoo JH, Koh IJ, Kang YG, Seong SC, Kim TK. Key factors in determining surgical timing of total knee arthroplasty in osteoarthritic patients: age, radiographic severity, and symptomatic severity. J Orthop Traumatol 2010;11-1:21-7.

- Choi WC, Lee S, Seong SC, Jung JH, Lee MC. Comparison between standard and high-flexion posterior-stabilised rotating-platform mobile-bearing total knee arthroplasties: a randomised controlled study. J Bone Joint Surg [Am] 2010;92-16:2634-42.
- Chowdry MM, Dagash H, Pierro A. A systematic review of the impact of volume of surgery and specialisation on patient outcome. British Journal of Surgery 2007;94(2):145-161.
- Cobb J. Joint registries electronic letter to the editor. J Bone Joint Surg [Br] Online 30th Jul 2010. http://web.jbjs.org.uk/cgi/eletters/92-B/6/749#5260
- Cobos R, Latorre A, Aizpuru F, Guenaga JI, Sarasqueta C, Escobar A, García L, Herrera-Espiñeira C Variability of indication criteria in knee and hip replacement: an observational study. BMC Musculoskelet Disord 2010;11:249.
- Culliford DJ, Maskell J, Beard DJ, Murray DW, Price AJ, Arden NK. Temporal trends in hip and knee replacement in the United Kingdom: 1991 to 2006. J Bone Joint Surg [Br] 2010;92:130-135.
- Cutler SJ, Ederer F. Maximum utilisation of the life-table method in analysing survival, J Chronic Dis 1958;8:699-712.
- Dalury DF, Tucker KK, Kelley TC. All-polyethylene tibial components in obese patients are associated with low failure at midterm follow-up. Clin Orthop Relat Res 2012;470(1):117-24.
- Danielson V. A Report on the PROMs programme for hip and knee replacement 2009 (Internal correspondence).
- Darzi A. High quality care for all: NHS Next Stage Review final report. NHS Next Stage Review. Department of Health 2008:1-92.
- Davidson D, Graves S, Tomkins A, *et al*. Australian Orthopaedic Association: annual report 2009: hip and knee arthroplasty September 1999 to December 2008. http://www.dmac.adelaide.edu.au/aoanjrr/documents/aoanjrrreport_2009.pdf
- Davies AP, Rating systems for total knee replacement. The Knee 2002;9:261-6.
- Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. BMJ 2010;340:doi:10.1136/bmj.c186.
- Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg [Br] 1998;80:63–69.

- Deehan D, Murray JD, Birdsall PD, Pinder IM. Quality of life after knee revision arthroplasty. Acta Orthopaedica 2006;77:761-766.
- Department of Health: Health and Social Care Information Centre. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_091815
- Department of Health: 2011-12 tariff information spreadsheet. 23rd March 2012. At http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyandGuidance/DH_124356
- Department of Health: Guidance on the routine collection of Patient Reported Outcome Measures (PROMs). February 2009 Vol. 2011: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_092647.
- Department of Health: Health and Social Care Information Centre. Patient Reported Outcome Measures (PROMs) report 2011.
- Dequeker J, Dieppe PA, eds. Disorders of bone cartilage and connective tissue. In: Klippel JH, Dieppe PA, eds. Rheumatology. 2nd ed. London: Mosby, 1998.
- Dervin GF, Carruthers C, Feibel RJ, Giachino AA, Kim PR, Thurston PR. Initial experience with the Oxford unicompartmental knee arthroplasty. J Arthroplasty 2011;26:192-7.
- Deshmukh RG, Hayes JH, Pinder IM. Does body weight influence outcome after arthroplasty? A 1-year analysis. J Arthroplasty 2002;17:315-9.
- Dowsey MM, Choong PFM. Early outcomes and complications following joint arthroplasty in obese patients: a review of the published reports. ANZ J Surg 2008;78:439-44.
- Dowsey MM, Liew D, Stoney JD, Choong PF. The impact of pre-operative obesity on weight change and outcome in total knee replacement. A prospective study of 529 consecutive patients. J Bone Joint Surg [Br] 2010;92-B:513-20.
- Drewett RF, Minns RJ, Sibly TF. Measuring outcome of total knee replacement using quality of life indices. Annals of the Royal College of Surgeons 1992;74:286-290.
- Dudley TE, Gioe TJ, Sinner P, Mehle S. Registry outcomes of unicompartmental knee arthroplasty revisions. Clin Orthop Rel Res 2008;466:1666-70.

- Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from the Swedish knee Arthroplasty Registry. J Bone Joint Surg [Br] 2001;83:339-44.
- Ehrlich GE. The rise of osteoarthritis. Bulletin of the World Health Organization 2003;81(6):630.
- Endres S. High-flexion *versus* conventional total knee arthroplasty: a 5-year study. J Orthop Surg (Hong Kong) 2011;19-2:226-9.
- Escobar A, Quintana JM, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. Osteoarthritis Cartilage 2007;15-3:273-80.
- Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Heath Related quality of life in total hip and total knee arthroplasty: A qualitative and systematic review of the literature. J Bone Joint Surg [Am] 2004;86(5):963-974.
- EuroQol Group. EQ5D: a standardised instrument for use as a measure of health outcome. http://www.euroqol.org/home.html
- Fairley J. The routine measurement of clinical outcomes – threat, opportunity or a waste of money? Royal Society of Medicine Open Section Meeting, 16 June 2008.
- Fehring TK, Odum SM, Griffin WL, Mason JB, McCoy TH. The obesity epidemic: its effect on total joint arthroplasty. J Arthroplasty 2007;22(6 supplement 2):71-6.
- Fehring TK, Odum SM, Masonis JL, Springer BD. Early failures in unicondylar arthroplasty. Orthopedics 2010;33:11-1.
- Fender D, Harper WM, Gregg PJ. The Trent arthroplasty register. Experiences with a hip register. J Bone Joint Surg [Br] 2000;82(7):44-6.
- Fitzgerald JD, Orav EJ, Lee TH, Marcantonio ER, Poss R, Goldman L, Mangione CM. Patient quality of life during the 12 months following joint replacement surgery. Arthritis and rheumatism 2004;51(1):100-9.
- Font-Rodriguez De, Scuderi GR, Insall JN. Survivorship of cemented total knee arthroplasty. Clin Orthop Relat Res 1997;345:79-86.
- Foran FR, Mont MA, Etienne G, Jones LC, Hungerford DS. The outcome of total knee arthroplasty in obese patients. J Bone Joint Surg [Am] 2004;86-A:1609-15.

- Fortin PR, Clarke AE, Joseph L, Liang MH, Tanzer M, Ferland D, Phillips C, Partridge AJ, Belisle P, Fossel AH, Mahomed NN, Sledge CB, Katz JN. Outcomes of total hip and knee replacement: Preoperative Functional Status Predicts Outcomes at six months after surgery. Arthritis and Rheumatism 1999;42-8:1722-28.
- Freeman MA, Pinskerova V. The movement of the normal tibio-femoral joint. J Biomech 2005;38:197-208.
- Furnes O, Espehaug B, Lie SA, Vollset SE, Engesaeter LB, Havelin LI. Failure mechanisms after unicompartmental and tricompartmental primary knee replacement with cement. J Bone Joint Surg [Am] 2007;89(3):519-25.
- Gardner J, Gloe J, Tatman P. Can this prosthesis be saved? Implant salvage attempts in infected primary TKA. Clin Orthop Relat Res 2011;469:970-976.
- Garellick G, Malchau H, Herberts P. Survival of hip replacements. A comparison of a randomised control trial and a registry. Clin Orthop Relat Res 2000;157-167.
- Garratt AM, Brealey S, Gillespie WJ. Patient assessed health instruments for the knee: a structured review. Rheumatology 2004;43:1414-1423.
- Ghomrawi HMK, Kane RL, Eberly LE, Bershadsky B. Patterns of functional improvement after revision knee arthroplasty. J Bone Joint Surg [Am] 2009;91:2838-2845.
- Gioe TJ, Killeen KK, Hoeffel DP, Bert JM, Comfort TK, Scheltema K, Mehle S, Grimm K. Analysis of unicompartmental knee arthroplasty in a community-based implant registry. Clin Orthop Relat Res 2003;(416):111-9.
- Göksan SB, Freeman MAR.One-Stage reimplantation for infected total knee arthroplasty. J Bone Joint Surg [Br] 1992;74-B:78-82.
- Goodfellow JW, O'Connor JJ, Murray DW. A critique of revision rate as an outcome measure: Reinterpretation of knee Joint Registry data. J Bone Joint Surg [Br] 2010;92-B:1628-31.
- Gooding CR, Masri BA, Duncan CP, Greidanus NV, Garbuz DS. Durable infection control and function with the PROSTALAC spacer in two-stage revision for infected knee arthroplasty. Clin Orthop Rel Res 2011;469:985-993.
- Gould, W. Interpreting logistic regression in all its forms. Stata Technical Bulletin 2000 sg124:19-29.

- Graves SE, Davidson D, Ingerson L, Ryan P, Griffith EC, McDermott BF, McElroy HJ, Pratt NL. The Australian Orthopaedic Association National Joint Replacement Registry. Med J Aust 2004;180(5 Suppl.):S31–4.
- Graves SE. The value of arthroplasty registry data, Acta Orthop 2010:81(1):8-9.
- Green GA. Understanding NSAIDS: from aspirin to COX-2. Clinical Cornerstones 2001;3:50-59.
- Greidanus NV, Peterson RC, Masri BA, Garbuz DS. Quality of life outcomes in revision *versus* primary total knee arthroplasty. J Arthroplasty 2011;26;615-620.
- Gunther T, Murray D, Miller R. Lateral unicompartmental knee arthroplasty with Oxford meniscal knee. Knee 1996;3:33-9.
- Hamilton WG, Ammeen D, Engh CA Jr, Engh GA. Learning curve with minimally invasive unicompartmental knee arthroplasty. J Arthroplasty 2010;25(5):735-40.
- Hanssen AD, Rand JA. A comparison of primary and revision total knee arthroplasty using the kinematic stabilizer prosthesis. J Bone Joint Surg [Am] 1988;70(4):491-499.
- Harcourt WGV, White SH, Jones P. Specificity of the Oxford Knee Status Questionnaire: the effect of disease of the hip or lumbar spine on patients' perception of knee disability. J Bone Joint Surg [Br] 2001;83B(3):345-7.
- Harmon JW, Tang DG, Gordon TA, Bowman HM, Choti MA, Kaufman HS, Bender JS, Duncan MD, Magnuson TH, Lillemoe KD, Cameron JL. Hospital volume can serve as a surrogate for surgeon volume for achieving excellent outcomes in colorectal resection. Ann Surg 1999;230:404–411.
- Hartley RC, Barton-Hanson NG, Finley R, Parkinson RW. Early patient outcomes after primary and revision total knee arthroplasty. J Bone Joint Surg [Br] 2002;84-B:994-999.
- Hawker G, Wright J, Coyte P, Paul J, Dittus R, Croxford R, Katz B, Bombardier C, Heck D, Freund D. Health related Quality of life after Knee Replacement. J Bone Joint Surg [Am] 1998;80(2):163-173.
- Havelin LI, Espehaug B, Vollset SE, Engesaeter LB. Early failures among 14,009 cemented and 1,326 uncemented prostheses for primary coxarthrosis. The Norwegian Arthroplasty Register, 1987–1992. Acta Orthop Scand 1994;65(1):1–6.

- Hervey SL, Purves HR, Guller U, Toth AP, Vail TP, Pietrobon R. Provider Volume of Total Knee Arthroplasties and Patient Outcomes in the HCUP-Nationwide Inpatient Sample. J Bone Joint Surg [Am] 2003;85-A(9):1775-83.
- Heyse TJ, Tibesku CO. Lateral unicompartmental knee arthroplasty: a review. Arch Orthop Trauma Surg 2010;130:1539-48.
- Horan FT. Joint registries. J Bone Joint Surg [Br]2010;92-B:749-50.
- Hospital Episode Statistics (HES). HES online: http://www.hesonline.org.uk/Ease/ContentServer?siteID=1937&categoryID=1295
- Hospital Episode Statistics. Patient Reported Outcome Measures report 2009-2010. HES online: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=295
- Hossein F, Patel S, Haddad FS. Midterm assessment of causes and results of revision total knee arthroplasty. Clin Orthop Relat Res 2010;468:1221-1228.
- Iannone F, Lapadula G. The pathophysiology of osteoarthritis. Aging Clin Exp Res 2003;15(5):364-72.
- Jacobs W, Anderson P, Limbeek J, Wymenga A. Mobile bearing vs fixed bearing prostheses for total knee arthroplasty for post-operative functional status in patients with osteoarthritis and rheumatoid arthritis. Cochrane Database Syst Rev 2004-2:CD003130.
- Jameson SS, Baker PN, Mason J, Porter ML, Deehan DJ, Reed MR. Independent predictors of revision following metal-on-metal hip resurfacing: a retrospective study using National Joint Registry data. J Bone Joint Surg [Br] 2012;94(6):746-54.
- Jämsen E, Stogiannidis I, Malmivaara A, Pajamäki J, Puolakka T. Outcome of prosthesis exchange for infected knee arthroplasty: the effect of treatment approach. A systematic review of the literature. Acta Orthop 2009;80:67-77.
- Janse AJ, Gemke RJ, Uiterwaal CS, Van der Tweel I, Kimpen JL, Sinnema G. Quality of Life: patients and doctors don't always agree: a meta-analysis. Journal of Clinical Epidemiology 2004;57:653-661.
- Johnston L, MacLennan G, McCormack K, Ramsay C, Walker A. The Knee Arthroplasty Trial (KAT) design features, baseline characteristics, and two-year functional outcomes

after alternative approaches to knee replacement. J Bone Joint Surg [Am] 2009;91-1:134-41.

- Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. Journal of Rheumatology 2000;27:1745–52.
- Judge A, Arden NK, Price A, Glyn-Jones S, Beard D, Carr AJ, Dawson J, Fitzpatrick R, Field RE. Assessing patients for joint replacement: Can pre-operative Oxford hip and knee scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection? J Bone Joint Surg [Br] 2011;93-B:1660-1664.
- Judge A, Chard J, Learmonth I, Dieppe P. The effects of surgical volumes and training centre status on outcomes following total joint replacement: analysis of the Hospital Episode Statistics for England. J Public Health (Oxf) 2006;28(2):116-24.
- Jones CA, Beaupre LA, Johnstaon DWC, Suarez-Almazor ME. Total joint arthroplasties: current concepts of patient outcomes after surgery. Clinics in Geriatric Medicine 2005;21:527-541.
- Kaplan EL, Meier P. Non parametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457-81.
- Katz JN, Mahomed NN, Baron JA, Barrett JA, Fossel AH, Creel AH, Wright J, Wright EA, Losina E. Association of hospital and surgeon procedure volume with patient-centered outcomes of total knee replacement in a population-based cohort of patients age 65 years and older. Arthritis Rheum. 2007;56(2):568-74.
- Katz JN, Barrett J, Mahomed NN, Baron JA, Wright RJ, Losina E. Association between hospital and surgeon procedure volume and the outcome of total knee replacement. J Bone Joint Surg [Am] 2004;86(9):1909-16.
- Kay P. Patient reported outcome measurement data (PROMS) British Orthopaedic News 2011;49:1-2.
- Keys GW, Ul-Abiddin Z, Toh, EM. Analysis of the first forty Oxford medial unicompartmental knee replacements from a small district hospital in UK. Knee 2004;11:375-7.

- Kim YH, Choi Y, Kim JS. Comparison of a standard and a gender-specific posterior cruciate-substituting high-flexion knee prosthesis: a prospective, randomised, short-term outcome study. J Bone Joint Surg [Am] 2010;92-10:1911-20.
- Kim HA, Kim S, Seo YI, Choi HJ, Seong SC, Song YW, Hunter D, Zhang Y. The epidemiology of total knee replacement in South Korea: national registry data. Rheumatology (Oxford) 2008;47:88-91.
- Kim J, Nelson CL, Lotke PA. Stiffness after total knee arthroplasty. J Bone Joint Surg [Am] 2004;86(7):1479-1484.
- Kline, R.B. Principles and Practice of Structural Equation Modelling. The Guilford Press , New York, London. 2003:pp427.
- Knutson K, Lindstrand A, Lidgren L. Survival of knee arthroplasties. A nation-wide multicentre investigation of 8000 cases. J Bone Joint Surg [Br] 1986;68(5):795–803.
- Kolling C, Simmen BR, Labek G, Goldhahn J. Key factors for a successful national arthroplasty register. J Bone Joint Surg [Br] 2007;89:1567-73.
- Koskinen E, Eskelinen A, Paavolainen P, Pulkkinen P, Remes V. Comparison of survival and cost-effectiveness between unicondylar arthroplasty and total knee arthroplasty in patients with primary osteoarthritis: a follow up study of 50493 knee replacements from the Finnish arthroplasty register. Acta Orthop 2008;79:499-507.
- Koskinen E, Paavolainen P, Ylinen P, Eskelinen A, Harilainen A, Sandelin J, Tallroth K, Remes V. Mid-term results for three contemporary total knee replacement designs - a comparative study of 104 patients with primary osteoarthritis. Scand J Surg 2010;99-4:250-5.
- Kretzer JP, Jakubowitz E, Reinders J, Lietz E, Moradi B, Hofmann K, Sonntag R. Wear analysis of unicondylar mobile bearing and fixed bearing knee systems: a knee simulator study. Acta Biomater 2011;7:710-15.
- Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of Primary and Revision Hip and Knee Arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg [Am] 2007;89(4):780-785.
- Labek G, Sekyra K, Pawelka W, Janda W, Stöckl B. Outcome and reproducibility of data concerning the Oxford unicompartmental knee arthroplasty: a structured literature review including arthroplasty registry data. Acta Orthop 2011;82(2):131-5.

- Laffer RR, Graber P, Ochsner PE, Zimmerli W. Outcome of prosthetic knee-associated infection: evaluation of 40 consecutive episodes at a single centre. Clin Microbiol Infect 2006;12:433-439.
- Laurencin CT, Zelicof SB, Scott RD, Ewald FC. Unicompartmental *versus* total knee arthroplasty in the same patient: a comparative study. Clin Orthop Relate Res 1991;273:151-56.
- Lavernia CJ, Guzman JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. J Arthroplasty 1995;10(2):133-40.
- Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, Heyse SP, Hirsch R, Hochberg MC, Hunder GG, Liang MH, Pillemer SR, Steen VD, Wolfe F. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis and Rheumatism. 1998;43:778-799.
- Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, Gabriel S, Hirsch R, Hochberg MC, Hunder GG, Jordan JM, Katz JN, Kremers HM, Wolfe F; National Arthritis Data Workgroup. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58:26-35.
- Lentino JR. Prosthetic joint infections: Bane of orthopaedists, challenge for infectious disease specialists. Clin Infect Dis 2003;36:1157-1161.
- Leone S, Borre S, D'Arminio Monforte A, Mordente G, Petrosillo N, Signore A, Venditti M, Viale P, Nicastri E, Lauria FN, Carosi G, Moroni M, Ippolito G, and the GISIG. Consensus document on controversial issues in the diagnosis and treatment of prosthetic joint infections. Int J Infect Dis 2010;14S4:S67-S77.
- Lewold S, Robertsson O, Knutson K, Lidgren L. Revision of unicompartmental knee arthroplasty: outcome in 1,135 cases from the Swedish Knee Arthroplasty study. Acta Orthop Scand 1998:69;469-74.
- Li MG, Yao F, Joss B, Ioppolo J, Nivbrant B, Wood D. Mobile vs. fixed bearing unicondylar knee arthroplasty: a randomised study on short term clinical outcomes and knee kinematics. Knee 2006;13:365-70.
- Lidgren L, Robertsson O, W-Dahl A. The Swedish Knee Arthroplasty Register: annual report 2009. http://www.knee.nko.se/english/online/uploadedFiles/113_SVK2009ENGL1.0.pdf

- Lingard EA, Katz JN, Wright EA, Sledge CB. Predicting the outcome of total knee arthroplasty. J Bone Joint Surg [Am] 2004;86-A-10:2179-86.
- Lisowski LA, van den Bekerom MP, Pilot P, van Dijk CN, Lisowski AE. Oxford Phase 3 unicompartmental knee arthroplasty: medium-term results of a minimally invasive surgical procedure. Knee Surg Sports Traumatol Arthrosc 2011;19:277-84.
- Lohmander LS, Gerhardsson de Verdier M, Rollof J, Nilsson PM, Engström G. Incidence of severe knee and hip osteoarthritis in relation to different measures of body mass: a population-based prospective cohort study. Annals of the Rheumatic Diseases 2009;68(4):490-6.
- Louie GH, Ward MM. Changes in the rates of joint surgery among patients with rheumatoid arthritis in California 1983-2007. Ann Rheum Dis 2007;69:868-71.
- Luring C, Kauper M, Bathis H, Perlick L, Beckmann J, Grifka J, Tingart M, Rath B. A five to seven year follow-up comparing computer-assisted vs freehand TKR with regard to clinical parameters. Int Orthop 2012;36(3):553-8.
- Lustig S, Elguindy A, Servien E, Fary C, Munini E, Demey G, Neyret P. 5- to 16-year follow-up of 54 consecutive lateral unicondylar knee arthroplasties with a fixed-all polyethylene bearing. J Arthroplasty 2011;26:1318-25.
- Lutzner J, Kasten P, Gunther KP, Kirschner S. Surgical options for patients with osteoarthritis of the knee. Nat Rev Rheumatol 2009;5:309-316.
- Lygre SH, Espehaug B, Havelin LI, Furnes O, Vollset SE. Pain and function in patients after primary unicompartmental and total knee arthroplasty. J Bone Joint Surg [Am] 2010;15-A:2890-7.
- Macmull S, Bartlett W, Miles J, Blunn GW, Pollock RC, Carrington RWJ, Skinner JA, Cannon SR, Briggs TWR. Custom-made hinged spacers in revision knee surgery for patients with infection, bone loss and instability. Knee 2010;17:403-406.
- Mancuso CA, Ranawat CS, Esdaile JM, Johanson NA, Charlson ME. Indications for total hip and knee arthroplasties: results of orthopaedic surveys. J Arthroplasty 1996;11:34-46.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Nat Cancer Inst 1959;22:719-48.

- Macario A, Schilling P, Rubio R, Bhalla A, Goodman S. What questions do patients undergoing lower extremity joint replacement surgery have? BMC Health Serv Res 2003;3-1:11.
- Malviya A, Bettinson K, Kurtz SM, Deehan DJ. When Do Patient-reported Assessments Peak after Revision Knee Arthroplasty? Clin Orthop Relat Res 2012;470:1728-1734.
- March LM, Cross MJ, Lapsley H, Brnabic AJ, Tribe KL, Bachmeier CJ, Courtenay BG, Brooks PM. Outcomes after hip or knee replacement surgery for osteoarthritis. A prospective cohort study comparing patients' quality of life before and after surgery with age-related population norms. The Medical Journal of Australia 1999;171(5):235-8.
- Marlow NE, Barraclough B, Collier NA, Dickinson IC, Fawcett J, Graham JC, Maddern GJ. Centralisation and the relationship between volume and outcome in knee arthroplasty procedures. ANZ J Surg 2010;80:234-41.
- Massoud SN, Hunter JB, Holdsworth BJ, Wallace WA, Juliusson R. Early femoral loosening of one design of cemented hip replacement. J Bone Joint Surg [Br] 1997;67-B:603–607.
- Mercier N, Wimsey S, Saragaglia D. Long-term clinical results of the Oxford medial unicompartmental knee arthroplasty. Int Orthop 2010;34:1137-43.
- Mittal Y, Fehring TK, Hanssen A, Marculescu C, Odum SM, Osmon D. Two-stage reimplantation for periprosthetic knee infection involving resistant organisms. J Bone Joint Surg [Am] 2007;89:1227-1231.
- Mow VC, Ratcliffe A, Poole AR. Cartilage and diarthrodial joints as paradigms for hierarchical materials and structures. Biomaterials 1992;13(2):67-97.
- Murray DW, Carr AJ, Bulstrode C. Survival analysis of joint replacements. J Bone Joint Surg [Br] 1993;75-B:697-704.
- Murray DW, Goodfellow JW, O'Connor JJ. The Oxford medial unicompartmental arthroplasty: a ten-year survival study. J Bone Joint Surg [Br] 1998;80:983-9.
- Najibi S, Iorio R, Surdam JW, Whang W, Appleby D, Healy WL. All-polyethylene and metal-backed tibial components in total knee arthroplasty: a matched pair analysis of functional outcome. J Arthroplasty 2003;18-7(Suppl 1):9-15.
- Namba RS, Paxton L, Fithian DC, Stone ML. Obesity and perioperative morbidity in total hip and knee arthroplasty patients. J Arthroplasty 2005;20:46-50.

- National Health Service (NHS): Better outcomes for children with congenital heart disease. February 16, 2011. NHS Specialised Services: http://www.specialisedservices.nhs.uk/news/view/25
- National Health Service Information Centre: http://www.ic.nhs.uk/webfiles/publications/003_Health_Lifestyles/OPAD12/Statistics_on_Obesity_Physical_Activity_and_Diet_England_2012.pdf
- National Health Service North Lincolnshire. Prior approval scheme 2009/10: http://www.northlincolnshire.nhs.uk/documents/download964.aspx.
- National Health Service Warwickshire. Commision policy statement: referral and surgical threshold for elective primary knee replacement surgery. April 2011: http://www.warwickshire.nhs.uk/CMSDocuments/b9ba5c0e-ff5a-425f-baf6-b323b08ce008.pdf
- National Institute for Clinical Excellence. NICE CG59 Osteoarthritis: The care and management of osteoarthritis in adults.
- National Institute for Clinical Excellence. NICE IPG230 Arthroscopic knee washout, with or without debridement, for the treatment of osteoarthritis.
- National Institute for Health Research (NIHR): National Institute for Health Research repository 2012: www.nihr.ac.uk/
- National Joint Registry for England and Wales. 9th Annual Report 2012: www.njrcentre.org.uk
- National Joint Registry for England and Wales. The research sub-committee: http://www.njrcentre.org.uk/njrcentre/Research/NJRPROMs/tabid/203/Default.aspx
- Nerhus TK, Heir S,Thornes E, Madsen JE, Ekeland A. Time-dependent improvement in functional outcome following LCS rotating platform knee replacement A prospective longitudinal study in 50 patients. Acta Orthop 2010;81(6):727–732.
- Newman J, Pydisetty RV, Ackroyd C. Unicompartmental or total knee replacement. The 15-year results of a prospective randomised controlled trial. J Bone Joint Surg [Br] 2009;91-B:52-7.
- Nizard RS, Biau D, Porcher R, Ravaud P, Bizot P, Hannouche D, Sedel L. A meta-analysis of patellar replacement in total knee arthroplasty. Clin Orthop Relat Res 2005-432:196-203.
- No authors listed. Euroqol (EQ5D) score: http://www.euroqol.org/home.html

- No authors listed. Annual report 2009: Hip and knee Arthroplasty September 1998 to December 2008. Australian Orthopaedic Association National Joint Replacement Register: http://www.dmac.adelaide.edu.au/aoanjrr/documents/aoanjrrreport_2009.pdf
- No authors listed. Danish Knee arthroplasty register: 2010 Annual report: http://www.dkar.dk/
- No authors listed. New Zealand Orthopaedic Association. The New Zealand Joint Registry. Eleven year report: January 1999 – December 2009 (2010): http://www.cdhb.govt.nz/njr/reports/A2D65CA3.pdf
- No authors listed. The Norwegian Arthroplasty Register: Annual report 2010: http://nrlweb.ihelse.net
- Noble PC, Conditt MA, Cook KF, Mathis KB. Patient expectations affect satisfaction with total knee arthroplasty. Clin Orthop Relat Res 2006;453:35-43.
- Norman-Taylor FH, Paler CR, Villar RN. Quality of life improvement compared after hip and knee replacement. J Bone Joint Surg [Br] 1996;78-B:74-7.
- Norton EC, Garfinkel SA, McQuay LJ, Heck DA, Wright JG, Dittus R, Lubitz RM. The effect of hospital volume on the in-hospital complication rate in knee replacement patients. Health Serv Res 1998;33(5 Pt1):1191-210.
- Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM. Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintainence organization. Arthritis and Rheumatism 1995;38:1134-1141.
- Pandit H, Jenkins C, Gill HS, Dodd CA, Murray DW. Minimally invasive Oxford phase 3 unicompartmental knee replacement: results of 1000 cases. J Bone Joint Surg [Br] 2011;93(2):198-204.
- Parkinson RW, Kay PR, Rawal A. A case for one-stage revision in infected total knee arthroplasty? Knee 2011;18:1-4.
- Parmet.S, Lynm C, Glass.RM. Osteoarthrotis of the knee. Journal of the American Medical Association 2003;289(8):1068.
- Pearse AJ, hooper GJ, Rothwell A, Frampton C. Survival and functional outcome after revsion of a unicompartmental to a total knee replacement. The New Zealand National Joint Registry. J Bone Joint Surg [Br] 2010;92-B:508-12.

- Pellegrini VD Jr, Marx C, Rankin EA, Batten JC, Frank CB, Atkinson D, van der Jagt DR. Position statement in support of National Joint Registries. J Bone Joint Surg [Am] 2009;91;2983.
- Pennington DW, Swienckowski JJ, Lutes WB, Drake GN. Lateral unicompartmental knee arthroplasty: survivorship and technical considerations at an average follow-up of 12.4 years. J Arthroplasty 2006;21:13-17.
- Petrie A. Statistics in orthopaedic papers. J Bone Joint Surg [Br] 2006;88-B:1121-36.
- Price AJ, Dodd CA, Svard UG, Murray DW. Oxford medial unicompartmental knee arthroplasty in patients younger and older than 60 years of age. J Bone Joint Surg [Br] 2005;87-B:1488-92.
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2010: http://www.R-project.org/
- Rajesekhar C, Das S, Smith A. Unicompartmental knee arthroplasty. 2- to 12-year results in a community hospital. J Bone Joint Surg [Br] 2004;86-B:983-5.
- Ranstam J, Kärrholm J, Pulkkinen P, Mäkelä K, Espehaug B, Pedersen AB, Mehnert F, Furnes O; NARA study group. Statistical analysis of arthroplasty data. I. Introduction and background. Acta Orthop 2011;82(3):253-7.
- Ranstam J, Kärrholm J, Pulkkinen P, Mäkelä K, Espehaug B, Pedersen AB, Mehnert F, Furnes O; NARA study group. Statistical analysis of arthroplasty data. II. Guidelines. Acta Orthop. 2011;82(3):258-67.
- Rawlins M. Harveian Oration. De testimonio: on the evidence for decisions about the use of therapeutic interventions. Lancet 2008;372:2152-61.
- Rissanen P, Aro S, Slatis P, Sintonen H, Paavolainen P. Health and Quality of life before and after hip or knee arthroplasty. J Arthroplasty 1995;10(2):169-175.
- Robertsson O, Lewold S, Knutson K, Lidgren L. The Swedish Knee Arthroplasty Project. Acta Orthop Scand 2000;71:7-18.
- Robertsson O, Knutson K, Lewold S, Lidgren L. The routine of surgical management reduces failure after unicompartmental knee arthroplasty. J Bone Joint Surg [Br] 2001;83:45-9.
- Robertsson O. Knee arthroplasty registers J Bone Joint Surg [Br] 2007;89(1):1-4.

- Robertsson O, Bizjajeva S, Fenstad AM, Furnes O, Lidgren L, Mehnert F, Odgaard A, Pedersen AB, Havelin LI. Knee arthroplasty in Denmark, Norway and Sweden. Acta Orthop 2010;86:82-89.
- Robertsson O, Dunbar M, Pehrsson T, Knutson K, Lidgren L. Patients satisfaction after knee arthroplasty: A report on 27,372 knees operated on between 1981 and 1995 in Sweden. Acta Orthop Scand 2000;71(3):262-267.
- Roos EM, Roos PH, Lohmander LS, Ekdahl, Beynnon BD. Knee injury and Osteoarthritis Outcome Score (KOOS): Development of a self-administered outcome measure. Journal of Orthopaedic and Sports Physical Therapy 1998;78(2):88-96.
- Rothwell PM, McDowell Z, Wong CK, Dorman PJ. Doctors and patients don't agree: cross sectional study of patient's and doctor's perceptions and assessment of disability in multiple sclerosis. British Medical Journal 1997;314:1580-1583.
- Sah AP, Scott RD. Lateral unicompartmental knee arthroplasty through a medial approach: study with an average five-year follow up. J Bone Joint Surg [Am] 2007;89-A:1948-54.
- Saleh KJ, Dykes DC, Tweedie RL, Mohamed K, Ravichandran A, Saleh RM, Gioe TJ, Heck DA. Functional outcome after total knee arthroplasty revision: a meta- analysis. J Arthroplasty 2002;17:967-977.
- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence based medicine: How to practice and teach EBM. London, Churchill Livingstone, 2000.
- Samson AJ, Mercer GE, Campbell DG. Total knee replacement in the morbidly obese: a literature review. ANZ J Surg 2010;80:595-99.
- Schulze Raestrup U, Smektala R. Are there relevant minimum procedure volumes in trauma and orthopedic surgery? Zentralbl Chir 2006;131(6):483-92.
- Scott CE, Howie CR, MacDonald D, Biant LC. Predicting dissatisfaction following total knee replacement: a prospective study of 1217 patients. J Bone Joint Surg [Br] 2010;92(9):1253-8.
- Scott IR, Stockley I, Getty CJM. Exchange arthroplasty for infected knee replacements. A new two-staged method. J Bone Joint Surg [Br] 1993;75-B:28-31.
- Scott RD. Lateral unicompartmental replacement: a road less travelled. Orthopaedics 2005;28:983-4.

- Sedrakyan A, Paxton EW, Phillips C, Namba R, Funahashi T, Barber T, Sculco T, Padgett D, Wright T, Marinac-Dabic D. The International Consortium of Orthopaedic Registries: Overview and Summary. J Bone Joint Surg [Am] 2011;93:Supplement 3.
- Shervin N, Rubash HE, Katz JN. Orthopaedic procedure volume and patient outcomes: a systematic literature review. Clin Orthop Relat Res 2007;457:35-41.
- Sibanda N, Copley LP, Lewsey JD, Borroff M, Gregg P, MacGregor AJ, Pickford M, Porter M, Tucker K, van der Meulen JHPLoS Med. Revision rates after primary hip and knee replacement in England between 2003 and 2006. PLOS Med 2008;5(9):e179.
- Smith SC, Cano S, Lamping DL, Staniszewska S, Browne J, Lewsey J, van der Meulen J, Ciarns J, Black N. Patient-Reported Outcome Measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific evidence. Final report to the Department of Health, December 2005.
- Sox HC, Greenfield S. Comparative Effectiveness Research: A report from the institute of medicine. Ann Intern Med 2009;151:203-205.
- Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. Lancet 2012;379(9822):1199-204.
- Spiegelhalter DJ. Funnel plots for comparing institutional performance. Statist Med 2005;24:1185-1202.
- Stuckenborg-Colsman C, Wirth CJ, Lazovic D, Wefer A. High Tibial Osteotomy *versus* unicompartmental joint replacement in unicompartmental knee joint osteoarthritis: 7-10 year follow up prospective randomised study. The Knee 2001;8(3):187-194.
- Suleiman LI, Ortega G, Ong'uti SK, Gonzalez DO, Tran DD, Onyike A, Turner PL, Fullum TM. Does BMI affect perioperative complications following total knee and hip arthroplasty? J Surg Res 2012;174(1):7-11.
- Svard UCG, Price AJ. Oxford medial unicompartmental knee arthroplasty: a survival analysis of an independent series. J Bone Joint Surg [Br] 2001;83-B:191-4.
- Symmons D, Mathers C, Pfleger B. Global burden of osteoarthritis in the year 2000. World Health Organisation. World Health Report 2002.
- Tanzer M, Smith K, Burnett S. Posterior-stabilised *versus* cruciate-retaining total knee arthroplasty: balancing the gap. J Arthroplasty 2002;17-7:813-9.

- The Audit Commission. Measuring quality: the patients's view of day surgery. London: The Audit Commission, 1991.
- The Healthcare Commission. State of Healthcare 2007. Improvements and challenges in services in England and Wales. Commission for Healthcare Audit and Inspection. Department of Health 2007:1-136.
- Therneau TM. Modelling Survival Data, Statistics for Biology and Heath. Springer, 2000.
- Whiteside LA, Peppers M, Nayfeh TA, Roy ME. Methicillin-resitant staphylococcus aureus in TKA treated with revision and direct intraarticular antibiotic infusion. Clin Orthop Rel Res 2011;469:26-33.
- Willis-Owen CA, Brust K, Alsop H, Miraldo M, Cobb JP. Unicondylar knee arthroplasty in the UK National Health Service: an analysis of candidacy, outcome and cost efficiency. Knee 2009;16;473-8.
- Winiarski R, Barth P, Lotke P. Total knee arthroplasty in morbidly obese patients. J Bone Joint Surg [Am] 1998;80:1770-4.
- World Health Organization. Obesity and overweight factsheet 2011: http://www.who.int/mediacentre/factsheets/fs311/en/index.html
- Wylde V, Blom A. The failure of survivorship. J Bone Joint Surg [Br] 2011;93(5):569-70.
- Wylde VV, Learmonth ID, Cavendish VJ. The Oxford hip score: The patient's perspective. Health and Quality of Life Outcomes 2005;3:12.
- Wylde V, Learmonth I, Potter A, Bettinson K, Lingard E. Patient-reported outcomes after fixed- *versus* mobile-bearing total knee replacement: a multi-centre randomised controlled trial using the Kinemax total knee replacement. J Bone Joint Surg [Br] 2008;90-9:1172-9.
- Yasunaga H, Tsuchiya K, Matsuyama Y, Ohe K. Analysis of factors affecting operating time, postoperative complications, and length of stay for total knee arthroplasty: nationwide web-based survey. J Orthop Sci 2009;14(1):10-6.
- Yeung E, Jackson M, Sexton S, Walter W, Zicat B, Walter W. The effect of obesity on the outcome of hip and knee arthroplasty. Int Orthop 2001;35(6):929-34.
- Zaninotto P, Head J, Stamatakis E, Wardle H, Mindell J. Trends in obesity among adults in England from 1993 to 2004 by age and social class and projections of prevalence to 2012. Journal of epidemiology and community health 2009;63(2):140-6.

Index

Glossary of abbreviations

Abbreviation	Term
AJR	Australian Joint Registry
AR	Annual Report
CER	Comparative effectiveness research
CI	Confidence Interval
DoH	Department of Health
DMARDs	Disease modifying anti-rheumatoid drugs
HES	Health Episode Statistics
HQIP	Health care Quality Improvement Programme
IC	Information Centre
KOOS	Knee injury and osteoarthritis outcome score
KSS	Knee Society Score
NCAPOP	National Clinical Audit and Patient Outcomes Programme
NARA	Nordic arthroplasty register association
NHS	National Health Service
NICE	National Institute for Clinical Excellence
NJR	National Joint Registry for England and Wales
NSAIDS	Non-steroidal anti-inflammatory drugs
NZJR	New Zealand Joint Registry
OKS	Oxford Knee Score
ONS	Office of National Statistics
PFA	Patello-femoral arthroplasty
PROMs	Patient Reported Outcome Measures
SAP	Scottish arthroplasty project
SEM	Structured equational modelling
TKR	Total Knee Replacement
UKR	Unicondylar Knee Replacement
VAS	Visual Analogue Scale