OUTCOMES EVALUATION IN KNEE ARTHROPLASTY FOR THE TREATMENT OF OSTEOARTHRITIS

by

MARY KATHLEEN SWEENEY

MSc., The University of British Columbia, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Pharmaceutical Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2010

ABSTRACT

Objectives: The primary objective was to evaluate differences in outcome between unicompartmental knee arthroplasty (UKA) and total knee arthroplasty (TKA) with respect to total WOMAC and OKS scores. In order to properly evaluate outcomes following both UKA and TKA valid and responsive instruments are needed. Therefore the second objective was to assess the construct validity and responsiveness of the SF-6D and the WOMAC derived HUI3 in patients undergoing either UKA or TKA. Lastly, the third objective was to evaluate any differences in the effect of socioeconomic status in outcomes following UKA compared to TKA.

Methods: A retrospective analysis was conducted to evaluate outcomes following knee arthroplasty on data collected between July 2000 and December 2008 on all patients requiring either UKA or TKA. Patients completed the WOMAC, OKS and SF-12 preoperatively and then subsequently at 3 and 6 months post-surgery. The Pearson correlation coefficient and intraclass correlation coefficient (ICC) were used to measure the correlation between the OKS, WOMAC, SF-6D and HUI3. Responsiveness was evaluated using a distribution based evaluation and an anchor-based method. A three level hierarchical linear model was used to model the total WOMAC scores and OKS as a function of group (UKA or TKA) and time. A similar model was used to investigate whether differences in total WOMAC or OKS between UKA and TKA occur across socioeconomic groups.

Results: The WOMAC derived HUI3 and SF-6D appear to be valid measures of HRQL in this patient sample. The SF-6D was not as responsive as the HUI3 but demonstrated a

good ability to identify patients who had improved. There were no significant differences in HRQL between UKA and TKA and socioeconomic status did not have a significant effect on HRQL following knee arthroplasty.

Conclusions: Both the HUI3 and SF-6D appeared to be valid and responsive measures in this patient sample. However, further study looking at the interchangeability of the WOMAC derived HUI3 with the HUI3 itself is needed. No differences in HRQL between UKA and TKA were noted and there were no differences in outcome between UKA and TKA across SES groups.

PREFACE

Chapter two in its entirety is currently under review for publication to the Journal of Bone and Joint Surgery under the title: An evaluation of the construct validity and responsiveness of generic utility measures in knee arthroplasty for the treatment of osteoarthritis. The candidate is first author of this manuscript which was co-authored by Dr. Carlo Marra and Dr. Larry Lynd, supervisor of the candidate, both of whom helped with the methodology and development of the primary hypothesis. Dr. Richard Kendall provided access to the data for which the analysis was based and Dr. Linda Li who was a member of the supervisory committee. The candidate's role in this manuscript involved development of the methodology, all data cleaning, statistical analysis and the study write-up.

Chapter three will be submitted for publication to the Journal of Orthopaedics and Related Research under the title: Comparison of HRQL between unicompartmental knee arthroplasty and total knee arthroplasty for the treatment of OA. The candidate is first author of this manuscript which is co-authored by Maja Grubisic, a statistician who provided statistical consultation and SAS programming in the analysis of the data. The other co-authors included Dr. Richard Kendall, an orthopaedic surgeon who provided surgical expertise and clinical knowledge, Dr. Larry Lynd, the candidate's supervisor who helped in the development of the conceptual framework and methodology of the study, and Drs. Carlo Marra and Linda Li, members of the supervisory committee who also provided comments regarding the manuscript during its preparation. The candidate was responsible for the statistical analysis, the development of the hypothesis, and wrote the final manuscript.

Chapter four will be submitted for publication to the Journal of Athroplasty under the title: The effect of socioeconomic status on outcomes following unicompartmental and total knee arthroplasty. The candidate is first author of this manuscript which is co-authored by Maja Grubisic, a statistician who provided statistical expertise and programming knowledge in the analysis of the data. Dr. Richard Kendall an orthopaedic surgeon who provided surgical input, Dr. Larry Lynd, the candidate's supervisor who proposed the use of the HLM model, and Drs. Carlo Marra and Linda Li, members of the supervisory committee. The candidate's role in this manuscript was the general direction of the work, the study hypothesis, data preparation, the statistical analysis, and the writing of the final manuscript.

Ethics approval was received from the University of British Columbia Research Ethics Board (#C04-0577) and research approval was received from the Richmond Health Services Research Advisory Committee for the patient data contained in all three manuscripts (appendix I).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank Dr. Larry Lynd for taking me on not only as a Master's student but as a player on his hockey team and a coach to his son's minor league team. I admire his patience, approachability, knowledge and ability to guide and teach not only me but all of his students. I would also like to thank Dr. Carlo Marra, for his light-hearted and sarcastic demeanour and his input and opinions with this project. And to my supervisory committee, Drs. Kendall, Li, Cairns and Greidanus for their insightful comments and encouragement throughout this project.

Last but not least, I want to thank my parents back in Ottawa who have been so supportive and understanding throughout this entire process. I could not have asked for any better role models.

CHAPTER 1

INTRODUCTION

1.1 OSTEOARTHRITIS

Osteoarthritis (OA) is the most common form of arthritis and it is estimated that three million Canadians currently suffer from the disease.[1] Osteoarthritis is a group of overlapping distinct diseases which have different etiologies but similar biologic, morphologic and clinical outcomes that affects not only the articular cartilage but involves the entire joint including the subchondral bone, ligaments, capsule, synovial membrane and periarticular muscles. It has been shown that OA is the result of a natural reaction of synovial joints to injury and therefore is the product of the impaired remodelling and repair process in the joint.[2]

The pathophysiology involves a combination of mechanical, cellular and biochemical processes that lead to changes in the composition and mechanical properties of the articular cartilage. The natural remodelling in healthy cartilage that occurs as the chondrocytes replace macromolecules lost during degredation is disrupted in osteoarthritis resulting in an abnormal repair process.[3] This impaired protective process causes the breakdown of the protein which serves as a cushion or protective layer between the bones causing the eventual loss of cartilage in the joint. Therefore previous injury to the joint, occupations that require repetition and strenuous movements, age, sex, and obesity are all risk factors.[4]

1.2 RISK FACTORS, TREATMENTS, AND DIAGNOSING OA

It is estimated that ninety percent of subjects 75 years of age and older have OA in any number of the weight bearing and non weight bearing joints e.g. digits of the hand, wrist knee, hip, spine, ankle shoulder. The number of people with OA is therefore expected to increase with the growth of the aging population.[5, 6]

1.2.1 The risk of developing OA associated with age and sex

Age is a strong risk factor for developing OA due to the large cumulative joint strain resulting from constant repair and change and as a result, OA is virtually imminent by age 80. After the age of 55 women are twice as likely to develop OA, particularly in the hand and knee due to the drop in oestrogen levels in post-menopausal women. In addition, moderate to severe symptoms are almost twice as prevalent in women as in men.

Over time age-related changes occur in the matrix composition leading to a decrease in chondrocyte function and responsiveness to stimuli. The cartilage diminishes in size and becomes unable to repair itself. Muscle strength diminishes as a result of decreased activity and fitness levels in the elderly. The decrease in activity level associated with aging results in changes in the matrix composition and eventual loss of function. However this effect can be arrested with the resumption of joint use which can help restore the joint composition to normal.[3]

1.2.2 Effect of increased body weight on developing OA

Appropriate body weight may be the single most important factor in preventing the onset of OA in the weight-bearing joints. It has been suggested that for every extra pound of body

weight, the knee carries an extra three pounds of extra stress.[7] A reduction in body weight has been shown to reduce the risk of developing OA because higher body weight can be associated with increased joint load per step and altered kinematics. However, excessive weight can be counterbalanced by mild activity levels which accompany a more sedentary lifestyle.[8]

1.2.3 Diagnosing OA

The clinical diagnostic criteria for OA proposed by the American College of Rheumatology consists largely of patient defined symptoms, the first being pain in the knee and at least three of the following: over the age of fifty, less than 30 minutes of morning stiffness, crepitus on active motion, bony tenderness, bony enlargement and no palpable warmth.[9]

Radiographs are also used in combination with patient defined symptoms to confirm the diagnosis using four central findings: joint-space narrowing between the bones, the formation of subchondral cysts (which occur because of micro fracture of the articular surface) and sclerosis which develops as new bone is laid down in the surrounding bone. Osteophytes can also form as disorganized new bone is produced at the margins of the joint.[10]

1.2.4 Operative and non operative treatment options for OA

Treatment options depend on the severity of symptoms but can broadly be subdivided in to non-operative and operative interventions. Knee replacement is a surgical procedure designed to diminish joint pain and improve function in patients with degenerative arthritis of the knee.[11] Typically, patients undergo knee arthroplasty after non-operative treatments such as activity modification, weight loss, bracing, anti-inflammatory medications, and/or knee joint injections

have failed to provide relief of arthritic symptoms.[12] There are two general types of knee replacements: total and unicompartmental knee arthroplasties.

In 2005 and 2006 there were over 40,000 hospitalizations for knee replacements in Canada and over the past ten years the number has increased by 69%.[13] In the United States on average 534, 000 knee replacements are performed every year with the rate of unicompartmental knee arthroplasty (UKA) growing three times faster than that of total knee arthroplasty.[14, 15, 16] This is likely due to the more recent introduction of the Oxford Bioment phase III prosthesis in 2004. However, eight percent of all knee arthroplasties performed in 2007 in the U.S. were unicompartmental yet the prevalence of OA affecting only one compartment is reported to range between 6% and 40% demonstrating its under-utilization.[17]

While many knee arthroplasties are successful for the short and longterm, revisions are often necessary for surgical failure. It is estimated that ten revisions are performed for every 100,000 Canadians from the general population, a revision being any complication resulting from a primary knee replacement.[13] Most of the reasons for failure of an implant are wear and osteolysis which can occur due to: an active lifestyle requiring deep flexion or kneeling, body mass index and body weight, limb alignment, gait mechanics and comorbidities. It has been shown that there are factors outside these patient level factors that contribute to failure as well such as issues surrounding the implant itself and surgical factor

1.3 TOTAL KNEE ARTHROPLASTY

A total knee arthroplasty (TKA) is a surgical procedure whereby diseased and non diseased portions of the knee joint are replaced with a mechanical joint. It is an effective

procedure yielding excellent long-term results yet, debate over prosthesis design and implant fixation still exists.[18] Most design concepts use metal on polyethylene for the bearing surfaces; however some use metal back patellar components with an optional cemented and un-cemented fixation.[19] Failure associated with material issues revolve predominantly around the type of metal and surface characteristics, the type of ultra-high-molecular weight polyethylene, modularity (modular metal-backed acetabular and tibial components) and type of implant fixation. In addition, insert thickness, locking mechanism, use of a mobile or rotating platform and the presence of an intact posterior cruciate ligament have also been shown to effect wear of the tibial components.[20]

Numerous studies have demonstrated no significant differences between types of prosthetic components in terms of survivorship and knee specific functional outcomes such as the Knee Society scores (KSS), and range of motion (ROM).[21-25] Specifically, a study performed by Gioe et al. analyzed 312 arthroplasties, 136 of which were all-polyethylene tibial components and 176 were of the fixed-bearing rotating platform design.[26] After a minimum of two years follow-up there was no significant difference between the mean postoperative ROM (110.9° and 109.1°, respectively; p=0.21), the mean KSS clinical score (90.4 and 88.2 points; p = 0.168) and ten patients required revision surgery (five in the all-polyethylene group and five in the fixed-bearing group). The causes of failure were deep infection, acute patellar fracture following a fall, symptomatic instability and aseptic loosening of the patellar component. There were no reported revisions due to polyethylene wear of the tibial component for either prosthetic design. However, it has been shown that the production of free floating polyethylene particles in any of the designs caused by wear can lead to osteolysis and eventual failure of the prosthesis.[27, 28] Jones et al. performed a detailed analysis on the failures caused by polyethylene wear in a series of 108

consecutive porous coated knee replacements.[29] Severe wear of the polyethylene tibial implant occurred in five cases (5%) which is a relatively high incidence compared to other studies that have reported this to be a rare occurrence.[30] For instance, Bryan and Rand reviewed a series of 5642 knee replacements and found an incidence of high density polyethylene wear of only 0.14% at an average follow-up of 5.5 years.[31]

Total knee arthroplasty also requires wide surgical exposure and is therefore a more invasive procedure resulting in a longer recovery period and inevitably an increased length of hospital stay relative to unicompartmental knee arthroplasty (UKA). Nevertheless, it has been shown to be reliable, efficacious and cost-effective for the treatment of end stage osteoarthritis of the knee and has been viewed as the gold standard for patients with arthritis of the knee.[32, 33] Most of these studies are observational and many of them use survivorship of the prosthesis as the main or only outcome[34-41] rather than patient focused health related quality of life (HRQL) outcomes.[42-45] A study performed by Wright et al. evaluated survivorship of the prosthesis, quality of life and patient satisfaction at 10 years post-surgery.[46] The probability of survival based on 407 patients (523 knees) who underwent primary total knee arthroplasty with the Kinemax prosthesis was 96.1% with revision for any reason as the end point. At the time of evaluation the mean Western Ontario and McMaster Osteoarthritis index (WOMAC) pain score was 87.9 and the mean WOMAC function score was 79.0 (with 100 being the best). The mean scores for physical functioning and physical role on the SF-36 were significantly lower than the general population ($p = 0.001$ and $p = 0.1$, respectively), but the mean score for general health perception was significantly higher ($p < 0.001$). In addition, patients were also generally very satisfied with the procedure with 74% stating that they would undergo surgery again and only 3.5% stating that they would not, and the rest were either unsure or did not respond.

1.4 UNICOMPARTMENTAL KNEE ARTHROPLASTY

Unicompartmental knee arthroplasty also known as partial knee replacement is an alternative to TKA, in which only the affected compartment is replaced, thus leaving the non diseased knee intact. Its reported advantages over TKA may include in: a shorter recovery, reduced hospital stay, less morbidity, fewer complications and easier revision surgery.[47] In approximately 70% of the cases of knee OA, only the medial compartment is affected and therefore re-establishing normal alignment often prevents the development of OA in the unaffected compartment, making UKA a feasible treatment.[48, 49]

The UKA prosthesis design requirements for the three main compartments of the knee i.e.: medial, lateral and patellofemoral, are very different and have yielded varying results demonstrating worse results in cases where OA affects the lateral compartment.[50] By preserving both the anterior cruciate ligament (ACL) and posterior cruciate ligament (PCL), UKA allows for the restoration of near-normal knee kinematics.

There are essentially two design approaches for unicompartmental prostheses both of which adhere to the crucial underlying idea that the anterior and posterior cruciate ligaments must be intact for normal knee function and implant longevity. Earlier studies (before 1985) did not use this in the inclusion criterion which is thought to have contributed to the high failure rate of unicompartmental procedures.[51, 52] The Oxford phase III prosthesis (Biomet®, Warsaw, Indiana), which was developed in 1999, offers a fully congruent meniscal bearing that is completely unconstrained allowing an evenly distributed allocation of weight throughout the joint, thus minimizing polyethylene wear as well as restoring knee kinematics.[53] Because polyethylene fragments which wear away from the insert into the joint have been shown to play

a role in aseptic loosening, thus increasing the need for revision, the Oxford prosthesis virtually eliminates this threat.[54]

The second unicompartmental prosthesis design uses a biconvex femoral component that articulates on a flat tibial plateau. The non congruous nature of the implant may lead to high polyethylene wear resulting in early failure. Limiting the excess of these free floating fragments is the main feature of the Oxford prosthesis.[55] Another inherent weakness of all unicompartmental implants is the need to use a relatively thin implant or, to sacrifice additional bone. Thin polyethylene cannot withstand the high pressures incurred because of small surface areas and it is known that incongruous surfaces require polyethylene at least 8mm thick.[56] Technique remains the most important aspect for this type of surgery as an overcorrection into valgus (outward angulation) alignment, may lead to development of OA in the unaffected compartment.[57]

Earlier studies have raised concerns that UKA is associated with an increased need for revision due to poorer survival of the prosthesis, improper patient selection and progression of arthritis in the untreated compartments due to malpositioning of the implant and malalignment of the joint.[58-62] Failure rates resulting in revision surgery were as high as 31% in a study performed by Mallory et al. with only 62% of procedures yielding 'acceptable' results postoperatively.[63] But since the introduction of UKA over forty years ago, many improvements have been made such as: better preoperative planning, proper patient selection and advanced implant design.[47, 51, 54, 64, 65] Price et al. for instance demonstrated a prosthetic survivorship of 93% at a minimum of 15 years follow-up using the Oxford (Biomet Ltd., Bridgend, UK) implant with polyethylene inserts of 6mm or less which translates into less sacrifice of bone.[53] In addition, the use of minimally invasive surgical techniques using a smaller incision have been reported to allow

patients more rapid recovery with less soft tissue trauma.[66, 67] As a result, patient's outcomes appear to have improved over time demonstrating comparability with total knee arthroplasty yet, there is a paucity of information available directly comparing the two procedures.[68, 69]

1.5 HEAD TO HEAD COMPARISONS BETWEEN UKA AND TKA

Several studies have compared UKA with TKA using different outcome measures ranging from survival of the prosthesis and cost to functional outcomes and patient satisfaction.[70, 71] Most of these studies have favoured UKA[72-80] yet these results, which have predominantly focused on survival of the prosthesis or one dimensional knee-specific outcomes, exclude relevant information relating to the overall health of the patient following knee replacement. In addition, several of these trials have compared groups with varying degrees of OA making it difficult to identify any true outcome differences.

A randomized control trial done by Newman et al. is to date the only well-controlled study comparing UKA and TKA.[81] A total of 102 patients were randomized, fifty of which received a UKA and 52 of which received a TKA. At a minimum of 15 years follow-up, 78% of patients in the UKA group had normal knee flexion versus only 38% in the TKA group. In the UKA group, 71.4% of patients reported excellent results on the Bristol Knee Score (BKS) compared to only 52.6% reporting the same in the TKA group.[82] Although this is the only study in the present literature to compare UKA and TKA using randomization into either group, the BKS and degree of knee flexion may not encompass all aspects of the overall improvement of the patients.

A study performed by Walton et al. attempted to account for patient-perceived change in outcome after either UKA or TKA using the Oxford Knee Score questionnaire (OKS) which assesses function and daily activities, the modified Gimby score which assesses levels of patient physical activity as the primary outcomes following surgery along with patient's levels of sports participation and working status.[83] The UKA group performed better on the OKS and modified Gimby score with mean scores at 1 year of 22.17 and 3.89, respectively, compared to 24.50 and 2.76, respectively, in the TKA group. Patients who underwent UKA were significantly more likely to increase or maintain their pre-operative level of sporting activity following surgery compared to TKA patients (P = 0.0003). The authors note that both groups were comparable with regards to age, sex and sports activity preoperatively however, the preoperative OKS and modified Gimby scores were significantly different between groups and therefore any conclusions drawn from this study were inherently flawed due to the lack of comparability at baseline. For this reason a comparison study properly evaluating patient-centered HRQL outcomes between UKA and TKA is needed. Moreover, valid outcome measures are needed that properly reflect the overall change in health following surgery specifically, generic HRQL instruments that can be used to facilitate their evaluation for cost-effectiveness analyses against other interventions.

1.6 GENERIC VERSUS DISEASE-SPECIFIC MEASURES

The goal of knee arthroplasty for both TKA and UKA is to improve and restore patient's quality of life. Patient self-assessed health related quality of life measurements, such as the Western Ontario and McMaster osteoarthritis index (WOMAC), have therefore become widely accepted for the evaluation of these procedures.[84] Consensus was reached among representatives

of the US Food and Drug Administration, European League Against Rheumatism, the World Health Organization/International League of Associations for Rheumatology, and the Group for the Respect of Ethics and Excellence in Science who were polled at the OMERACT III conference, recommending the use of the WOMAC index for the assessment, in phase III clinical trials, for hip, knee and hand osteoarthritis.[85] However since then, there have been few published comparisons by independent research groups. The Oxford Knee Score (OKS) is another example of a patient self-assessed questionnaire used in this study that is specific to dimensions of health related to knee pain, function and daily tasks and has been validated for use in the evaluation of HRQL following knee arthroplasty.[86] Alternatively, generic questionnaires such as the SF-12 measure the general health of the target population and are used across different health states. It has been shown that general instruments are better able to account for complications and conditions unrelated to the disease of interest.[87] Osteoarthritis is more prevalent in the elderly population and many sufferers have co-morbid conditions demonstrating the importance of a more complete view of HRQL in this patient group. For this reason, the evaluation of the validity and responsiveness of generic instruments is specifically relevant in the assessment of treatment effectiveness following knee arthroplasty.

Alternatively, the HUI3 and SF-6D are self-administered instruments that use population-derived weights to assign utility scores to specific health states that are anchored at 0 (death) and 1 (perfect health) and both of which have been used in rheumatology because they are easy to administer and appear to cover the health domains relevant to arthritis.[88] Utilities are most often used in economic evaluations and therefore are more commonly accepted generic measures of HRQL at least from a payer's point of view. In light of funding shortages, these preference weighted utility measures may therefore be used in the assessment of the economic benefit of

performing UKA over TKA, and to see if one offers a more cost-effective alternative by creating a higher quality of life for less cost.[89] Although the validity of generic utility measures has been studied in other disease states the validity and responsiveness of the SF-6D and HUI3 has yet to be evaluated in knee arthroplasty for the treatment of OA.[90, 91]

1.7 EFFECTS OF SOCIECONOMIC STATUS ON OUTCOMES FOLLOWING KNEE ARTHROPLASTY

Several explanatory variables have been identified that may be able to help predict HRQL outcomes following knee arthroplasty. Preoperative pain and physical function based on WOMAC scores have been shown to be strong predictors of outcome at 6 months, 1 year and two years in TKA.[92] Co-morbid conditions, BMI, age and mental health status are also highly correlated with outcomes following TKA.[93] Other patient factors such as: gender, type of arthritis, and patient expectations have also been correlated with outcome.[94, 95] However, information on the effects of socioeconomic status (SES) on patient outcomes following TKA is limited and there are currently no studies that have evaluated these effects on patient HRQL following UKA specifically or on the effects of SES on outcomes between patients undergoing UKA or TKA. Young et al. demonstrated that patients in lower socioeconomic groups were twice as likely to require total knee arthroplasty and were less likely to receive the surgery.[96] Furthermore, the effect of SES on the outcome of surgical interventions has been studied in other disease states. For example, lower socioeconomic status groups have been shown to have worse outcomes in other disease states requiring surgery such as liver and renal transplantation.[97, 98] A prospective multi-centered observational study conducted by Davis et al. showed that patients with lower income appeared to have a greater need for TKA demonstrating worse preoperative

disease-specific outcome scores yet these patients were able to compensate and obtain similar outcomes post-operatively.[99] The lone exception was the pain dimension score of the disease-specific questionnaire at 12 months.

1.8 RESEARCH NEEDS AND STUDY JUSTIFICATION

With the use of HRQL tools as a common outcome measure following knee arthroplasty there is still limited consensus regarding the use of generic utility scores in the evaluation of HRQL outcomes following knee replacement. The identification of appropriate instruments for evaluating the effectiveness of both UKA and TKA that are reflective of the change in health of patients following these procedures is therefore of great importance.

Given the large proportion of patients suffering from OA affecting the medial side of the knee and the less invasive nature of UKA along with the introduction of new prostheses design concepts, proper patient inclusion criteria and advanced surgical knowledge and experience, UKA could potentially be a comparable procedure to TKA yet, few comparison studies using clinical outcomes have been published that properly reflect these improvements.

Several patient level factors have been identified which help with expectations following total knee arthroplasty. Preoperative function for example has been shown to be one of the strongest determinants of outcome following total knee arthroplasty. Yet, little study has been undertaken regarding the effects of socioeconomic status in patients undergoing UKA and specifically in the identification of differences in outcome that may exist between UKA and TKA patients.

1.9 STUDY HYPOTHESIS AND RESEARCH OBJECTIVES

The aim of this study was to determine if there is any significant difference in HRQL between patients undergoing UKA compared to those undergoing TKA for the treatment of OA. The primary hypothesis of this study was that no significant difference would exist between the two treatment groups therefore suggesting that UKA is a viable option to TKA for patients suffering from OA affecting one compartment of the knee. Prior to the testing of this hypothesis, it was necessary to evaluate the HRQL and demographic characteristics of the two groups to ensure comparability at baseline and to assess the discriminative ability and responsiveness of the generic utility measurement instrument for the evaluation of the treatment effects on this sample following knee arthroplasty.

The **first** objective of this study was therefore to assess the validity and responsiveness of the SF-12 derived SF-6D utility and the WOMAC derived HUI3 in patients undergoing either UKA or TKA.

The **second** and primary objective was to evaluate differences in outcome between UKA and TKA with respect to total WOMAC scores and OKS after adjusting for severity of OA which was defined by pre-operative HRQL scores. The a priori hypothesis was that no significant difference would exist between patients that underwent UKA compared to TKA.

The **third** objective was to evaluate any differences in the effect of socioeconomic status on outcomes following UKA compared to TKA.

1.10 SUMMARY

Osteoarthritis of the knee is a common condition causing debilitation, pain and lack of mobility in patients suffering from the disease. Many factors are thought to contribute to its development and its prevalence is expected to increase significantly in Canada and the United States over the next twenty years. When non-surgical treatments such as medications, hyaline injections and activity modification fail to improve the quality of life, knee arthroplasty has been shown to be an effective treatment for end-stage osteoarthritis of the knee.

Two types of knee replacements have been used: unicompartmental or partial and total knee arthroplasty. Although total knee arthroplasty has been shown to be a successful treatment for osteoarthritis of the knee, technological advancements, surgical expertise and proper patient selection in unicompartmental knee replacement have offered long term durability comparable to that of total knee arthroplasty. This study therefore looks to determine if unicompartmental knee arthroplasty offers greater improvement in quality of life as compared to total knee replacement. We also hope to improve the knowledge pertaining to patient characteristics in order to provide physicians with the ability to use these as indicators when choosing either unicompartmental or total knee arthroplasty as a treatment for osteoarthritis. This will better equip clinicians when faced with the potential uncertainty surrounding treatment options for patients, based on age, sex, and socioeconomic status, with osteoarthritis of the knee.

Patient self-assessment tools have become increasingly used in the valuation of outcomes following knee arthroplasty and are therefore commonly used to evaluate health related quality of life. Pre-existing regression models will be used to transform the WOMAC and SF-12 into HUI3 and SF-6D utility scores respectively. These preference weighted health related utility measures may therefore be used in the assessment of the economic benefit of performing

unicompartmental arthroplasty over total knee arthroplasty, and to see if one offers a dominant alternative by creating a higher quality of life for less cost. However controversy still surrounds the use of generic utilities such as the SF-6D and HUI3 which strengths include applicability over a broad range of diseases and accountability for co-morbidities unrelated to the disease being treated, whereas condition-specific measures (AQL-5D and VisQoL) are thought to be more responsive to changes in health more closely related to the disease.

CHAPTER 2

AN EVALUATION OF THE CONSTRUCT VALIDITY AND RESPONSIVENESS OF GENERIC UTILITY MEASURES IN KNEE ARTHROPLASTY FOR THE TREATMENT OF OSTEOARTHRITIS

2.1 FORWARD

This chapter is currently under review for publication, under the same title, in the Journal of Bone and Joint Surgery. The candidate is first author of this manuscript which was co-authored by Dr. Carlo Marra and Dr. Larry Lynd, supervisor of the candidate, both of whom helped with the methodology and development of the primary hypothesis. Dr. Richard Kendall provided access to the data for which the analysis was based and Dr. Linda Li who was a member of the supervisory committee.

The candidate's role in this manuscript involved development of the methodology, all data cleaning, statistical analysis and the study write-up.

2.2 INTRODUCTION

The measurement of self-reported health-related quality of life (HRQL) has become widely accepted and has been used in the evaluation of the effectiveness of treatments in many chronic diseases.[84] The Western Ontario and McMaster Osteoarthritis (WOMAC) index is commonly used for this purpose in patients with osteoarthritis. In contrast to the WOMAC which measures disease specific items, the Short Form 12 (SF-12) is a generic tool that has been shown to be valid and reliable over a broad range of illnesses.[100]

These tools are examples of psychometric instruments that offer a straightforward way of measuring different dimensions or attributes of health using multi-item scales.[101] Alternatively, preference based measures, which are most commonly used in economic evaluations, use a single index score that is reflective of society's preference for a given health status. In general there are two ways that preference-based health-state utilities can be elicited: 1) directly, in which the patient is asked either how many days, months, years they would be willing to give up in exchange for perfect health (time trade off) or the magnitude of the risk (i.e. probability) of instantaneous death that they would be willing to take in exchange for achieving perfect health (standard gamble); and 2) indirectly, using multi-attribute utility instruments such as the Health Utilities Index (HUI3) and Short Form 6D (SF-6D) which are examples used in the present study and can be derived from the WOMAC and SF-12, respectively.[102]

The HUI3 and SF-6D are self-administered instruments that use population-derived weights to assign utility scores to specific health states which range from 0 (death) to 1 (perfect health). The HUI3 can theoretically describe 972,000 different preference weighted health states with possible utilities ranging from -0.36 (i.e. the worst possible health state that is worse than death) to 1.00 (perfect health).[103] Recently, a regression model was developed to map WOMAC scores onto the HUI3.[104] Similarly, the SF-12 can be used to derive the corresponding SF-6D utility score with 1.00 representing full health and 0 representing death; however, 0.30 is the worst possible living health state that can be experienced by a person using this scale.[105] While disease-specific instruments are able to capture minor changes in health relevant to the disease in question, comparisons cannot be made across different conditions.[106] Thus, one advantage to using generic instruments is that they combine different domains of health into a single index and thus can be used to compare across diseases.

Although the validity of generic utility measures has been evaluated in other disease states the validity and responsiveness of the SF-6D and HUI3 has yet to be evaluated in knee arthroplasty.[90, 91] Therefore, the objective of this study was to evaluate the construct validity and responsiveness of both the SF-12 derived SF-6D and the WOMAC derived HUI3 utilities following knee arthroplasty. The study consisted of three specific aims: 1) to evaluate the differences between health state utilities derived using the HUI3 and the SF-6D; 2) to evaluate and compare the responsiveness of the instruments to change; and 3) to assess the discriminative properties of the HUI3 and SF-6D across subjects who underwent either unicompartmental knee arthroplasty (UKA) or total knee arthroplasty (TKA).

2.3 PATIENTS AND METHODS

A retrospective analysis was conducted on data collected on all consenting patients undergoing either a UKA or TKA between July 2000 and November 2008 at Richmond General Hospital. The WOMAC, Oxford Knee Score (OKS) and SF-12 were completed preoperatively, and at 3 and 6 months following surgery. Five surgeons implanted all UKAs (Biomet® Oxford implant, Warsaw, Indiana) and TKAs (Advance® Wright Medical implant Memphis, Tennessee). Ethics approval was received from the University of British Columbia Research Ethics Board (#C04-0577) and research approval was received from the Richmond Health Services Research Advisory Committee.

The validity and responsiveness of the HUI3 and SF-6D were evaluated relative to the WOMAC and the OKS. The WOMAC is a 24-item osteoarthritis specific instrument that was originally designed for use in clinical trials in patients with OA of the knee or hip. It has been previously tested for validity, reliability and responsiveness, and is comprised of three domains:

pain (5 questions), stiffness (2 questions) and physical function (17 questions). The Likert 3.1 version was used in this study, with each domain score calculated by summing the coded responses corresponding to the domain specific questions, and a total score calculated by summing the scores for each domain. The scores for each item ranged from 0 (full health) to 4 (extreme disability).[107-109]

The OKS is a 12-item questionnaire that has been previously validated and is recommended for the evaluation of HRQL following knee arthroplasty.[110] A single score that ranges from 12 (full health) to 60 (extreme disability) is derived from patient's self-reported responses to questions relating to pain, knee function and day to day tasks and activities.[86]

Lastly, the Medical Outcomes Study 12 Item Short Form (SF-12) is a generic instrument designed to measure the general health of the target population. It consists of 12 questions encompassing seven domains: general health, physical function, role limitations, pain, mental health, vitality and social functioning.[111] It is comprised of both a mental component score (MCS) and a physical component score (PCS) that range from 0 (extreme disability) to 100 (full health).

2.3.1 Statistical analysis

Demographic variables were compared between patients who had complete data at each time point and those with incomplete data (missing questionnaires, age and or sex) using t-tests for continuous data and Pearson χ^2 test for categorical data. Patients receiving a UKA or TKA were grouped together under the assumption that the validity, responsiveness, and discriminative ability of each instrument is independent of the type of prosthesis implanted. For patients with bilateral knee replacements, only the first procedure was evaluated. Patients for whom there

were four or more missing responses for the SF-12 scores were excluded from the analysis of that instrument. Where less than four responses were missing, the missing responses were imputed using the method proposed by Perneger et al.[112] When imputing the PCS and MCS scores, the mean score of the missing question from the remaining sample was used in its place. For the WOMAC, if two or more pain items, both stiffness items, or four or more function items were missing, the patient was excluded from the analysis of that subscale. As recommended in the WOMAC user guide, where one pain, one stiffness, or up to three function items were omitted, the mean of the remaining responses for the specific subscale was used.[113] A similar method was used for the Oxford knee questionnaire in which the mean value representing all the other responses was used when two or fewer questions were omitted. If responses to more than two questions were omitted, the patient was excluded from that analysis.[114] All analyses were performed using SPSS 17.0 (SPSS Inc. Chicago, Illinois).

2.3.2 Utility derivation

Grootendorst et al. evaluated four separate regression models for deriving a HUI3 utility score from the WOMAC.[104] The model that best fit the data consisted of: WOMAC subscales with their interactions, age, gender, OA duration and their second order terms. In a separate study evaluating the validity of these models, it was found that the WOMAC subscales with their interactions and their square terms as well as age and sex demonstrated the lowest mean absolute error; therefore, we used this model to derive the HUI3 utility scores for this study.[115] The algorithm developed by Brazier et al. was used to derive SF-6D utilities from the SF-12.[116]

2.3.3 Construct validity

To evaluate the construct validity of the SF-6D and HUI3, we used the results from each questionnaire at baseline, and three and six month's follow-up. The correlation between the OKS, WOMAC, SF-6D and HUI3 was measured using the Pearson correlation coefficient. Given that the HUI3 and SF-6D both measure health state utilities (i.e. theoretically the same construct), these two instruments were compared using the intra-class correlation coefficient (ICC). This was done using a 2-way mixed effects model such that the instrument effect was fixed and the subject effect was random.[117] A correlation coefficient for both the ICC and the Pearson correlation greater than 0.5 or less than -0.5 were deemed to be strong and 0.30 to 0.49 or -0.49 to -0.30 were considered moderate. Instruments with correlation coefficients between 0.30 and -0.30 were considered weakly correlated.[118]

2.3.4 Responsiveness

The ability to detect clinically meaningful change (i.e. change that is meaningful to the patient) is important in evaluating the sensitivity of an instrument.[119] The responsiveness of the SF-6D and HUI3 in patients undergoing knee arthroplasty for the treatment of OA at three and six months was therefore assessed using two strategies: 1) using a distribution based evaluation of Cohen's effect size (ES)[120] and the standardized response mean (SRM);[121] and, 2) using a flexible dichotomous regression model to represent the anchor-based method, which uses an external criteria as a comparison to assess the discriminative ability of the SF-6D and HUI3 utility scores.[92]

In order to evaluate the magnitude of the treatment effect captured by the SF-6D and HUI3 utility scores, the effect size (*d*) was calculated as:

$$d = \frac{\overline{X}_1 - \overline{X}_2}{SD_{Baseline}}$$

where $\overline{X_1}$ represents the mean score at baseline $\overline{X_2}$ the mean score at three or six months, and the denominator is the standard deviation (SD) of the mean baseline scores. Following Cohen's recommendations, we considered effect sizes of 0.20 – 0.49 to be small, 0.50-0.79 to be moderate and ≥ 0.80 to be large.[122] The effect size of the SF-6D and HUI3 were then compared to that of the other disease-specific questionnaires at three and six months post-surgery. The mean change in the SF-6D score was then compared to that of the HUI3 utility scores using the t-test at 3 and 6 months.

Given the known effectiveness of knee arthroplasty for the treatment of osteoarthritis, it was hypothesized that at six months all patients would demonstrate some degree of improvement. We therefore calculated the standardized response mean (SRM), which is defined as the ratio of mean change to the standard deviation of the change scores in patients reporting change and allows for direct comparison of each instrument.[123] Thus, the

$$SRM = \frac{\overline{X}_1 - \overline{X}_2}{SD_{change}}$$

where $\overline{X_1}$ represents mean SF-6D and HUI3 scores at baseline and $\overline{X_2}$ the mean scores at three and six months post-operatively. The standardized response means were calculated for each instrument in all patients at 3 and 6 months follow-up. Non-parametric bootstrapping was used to generate 95% confidence intervals for both the ES and SRM[124]

Although a valid minimal clinically important difference (MCID) for the OKS following knee replacement has not been established, it has been recommended that half of the standard

deviation of the mean change score can be used.[125] In this sample, this translated into approximately a ten percent change in overall OKS score at six months. We therefore classified patients as improved using cut points of > 10%, >20%, and >30% improvement in OKS, while those below each threshold were classified as not improved, meaning they were either unchanged or got worse. Given that a 20% change in WOMAC pain scores has been validated as a clinically meaningful threshold, these cut-points were then plotted against the percent change WOMAC pain scores.[126] In other words, for any given percent change in scores, the graph illustrates the proportion of patients with that difference whose disease activity following knee arthroplasty is expected to be improved or not improved, according to the OKS at each cut-point.

The WOMAC and OKS cut points were then used as the external predictors to evaluate the discriminative ability of the SF-6D and HUI3. This was done in order to answer more clinically relevant questions such as, what is the magnitude of change in the SF-6D and HUI3 utility scores necessary to ensure a high probability that it reflects a clinically meaningful improvement. Subjects were thus divided into those whose WOMAC pain scores improved >20%, from baseline to 6 months following surgery and those whose WOMAC pain scores did not change (i.e. ≤ 20% change in the WOMAC pain scores). Given that the HUI3 was derived from the WOMAC and thus it could not be used as the external predictor a similar plot was created for the HUI3 using the OKS as the external predictor of change.

2.4 RESULTS

Of the 1110 patients that underwent knee arthroplasty between July 2000 to November 2008, 26 (2%) of knee replacements were second procedures and were therefore not included, two patients had no HRQL scores at any point and 347 (31%) were missing surgical information

which consisted of either the type of procedure or the surgeon who performed the surgery and were therefore excluded (Figure 2-1). Therefore, Seven hundred and thirty-seven patients were included in the analysis with a mean age of 70.5 years (SD 9.20), 62% of which were female. Two hundred and eighty-one (38%) of patients completed the questionnaires only at the preoperative assessment, 94 (13%) completed at least one questionnaire strictly at baseline and 3 months, 118 (16%) had follow-up assessments at 0 and 6 months and 244 (33%) of patients completed at least one questionnaire at all three assessment times. The WOMAC function score and SF-6D were significantly different at baseline between patients with data at only 3 or 6 months (n = 556) and those with all data at both follow-up assessments (n = 181) (Table 2-1). The SF-6D was strongly correlated with OKS, total WOMAC and function domain scores at each time point, but poorly correlated with the WOMAC stiffness score (Table 2-2). The HUI3 was strongly correlated with the total WOMAC, pain and function scores at baseline, 3 and 6 months follow-up. Both the HUI3 and SF-6D were strongly correlated with OKS (r = 0.66 and 0.62, respectively). The agreement between the HUI3 and SF-6D utilities was also strong at all time points (ICC = 0.59, 0.65 and 0.65 preoperative, 3 and 6 months, respectively).

2.4.1 Responsiveness

At least some improvement over time was detected by the total scores for all the measures (Table 2-3), with the exception of the SF-12 MCS for which there was little or no change in the mean scores from baseline following surgery. The mean change scores of the HUI3 and SF-6D utilities were not significantly different at three months (0.22 and 0.13, respectively; p = 0.090) and six months (0.24 and 0.15, respectively; p = 0.44). However, the mean absolute scores at the preoperative assessment (0.46 and 0.59 respectively; p < 0.001) at

the 3 month assessment (0.69 and 0.74 respectively; $p = 0.007$) and at 6 months (0.71 and 0.75 respectively; $p < 0.001$) were all significantly different between utilities. In general, the ESs and SRMs of the disease-specific measures (i.e. WOMAC and OKS) were larger than those of the generic health status and health utility scores (PCS, MCS and SF-6D). The MCS did not appear to be responsive to knee arthroplasty (i.e. ES and or SRM were less than one) but the ES and SRM of all instruments did increase from three to six months follow-up (Table 2-4).

The OKS, total WOMAC, WOMAC pain and the HUI3 were the most responsive based on the magnitude of the ES and SRM. Although the SF-6D was not as responsive as the HUI3 or the disease specific measures (WOMAC and OKS), it did appear to capture the change in HRQL resulting from knee arthroplasty.

2.4.2 Dichotomous response model

In determining an appropriate cut-off for the OKS as an external criteria for characterizing patients as either improved or not improved, it appeared that the 20 percent cut-off was able to better discriminate between those patients whose WOMAC scores (external predictor) improved or did not improve. For example, in Figure 2-2, Panel A, there was a fifty percent probability (0.5 on the y-axis) of being classified as improved given a 20 percent improvement (-20 on the x-axis) on the WOMAC pain score and an 85 percent probability (0.85 on the y-axis) of being classified as not improved given 20 percent (20 on the x-axis) deterioration on the WOMAC pain score. Alternatively, in Figure 2-2, Panel B, (10% cut-off for the OKS) there was a 65 percent probability of improvement (0.65 on the y-axis) given a 20 percent improvement on the WOMAC pain score and a 70 percent chance of being classified as not improved given 20 percent deterioration on the WOMAC pain score. For the 30% OKS cut-

off, a 20% improvement on the WOMAC pain score translated into only a 20 percent chance of being classified as improved on the OKS (graph not shown). Therefore, the 20% cut-off for the OKS is better able to discriminate between patients who improved and did not improve whereas the 10% cut-off is slightly more discriminative with only those patients who have improved. For this reason a 20% change in both the OKS and the WOMAC were used to define improvement. Thus, patients whose scores improved by at least 20% were classified as 'improved', while those whose scores did not improve by at least 20% were defined as 'not improved'.

In terms of the discriminative ability of the SF-6D, if a patient improved on the SF-6D then it was very likely that they had in fact improved according to the external predictor. For example, a 0.2 point change on the SF-6D (x-axis) was associated with almost a 95 percent probability of improvement according to the WOMAC pain score and the OKS (Figures 2-3, Panel A and B, respectively). Conversely, if a patient deteriorated by -0.2 points on the SF-6D, there was approximately a 65 percent chance that they had improved according to the WOMAC pain 20% cut-off (Figure 2-3, Panel A) and an 80 percent chance they had improved according to the OKS 20% cut-off (Figure 2-3, Panel B). An improvement of 0.2 on the HUI3 was associated with an 80 percent likelihood of improvement on the OKS and a -0.2 change on the HUI3 translated into a 95 percent chance of no improvement on the OKS (Figure 2-4).

2.5 DISCUSSION

It has been suggested that condition-specific HRQL measures are likely to be more responsive yet controversy still remains surrounding the use of generic tools in the evaluation of outcomes associated with knee arthroplasty for the treatment of osteoarthritis.[127] Therefore, the validation of generic utility measures is of particular importance when dealing with chronic diseases that

have virtually no mortality in the short term. This is the first study to show that both the SF-6D and the WOMAC derived HUI3 can be used to evaluate HRQL following knee arthroplasty. Both the SF-6D and HUI3 demonstrated good construct validity and were strongly correlated with two disease-specific questionnaires (WOMAC and OKS). The HUI3 was highly responsive to change and able to clearly discriminate between patients who improved and those that did not, based on the external criteria applied. Although the SF-6D was not as responsive as the WOMAC derived HUI3, it had good discriminative properties. Therefore, if a patient reported no change or worsening on the SF-6D there was only a 25-35 percent probability that they did in fact improve based on the external criteria. Yet if a patient improved by at least 0.2 on the SF-6D one can be confident (i.e. a 95 percent probability) that the individual did in fact improve based on the external criteria. Although the WOMAC derived HUI3 and SF-6D theoretically measure the same construct (i.e. health utility), there were differences in the mean scores, which were significant and the mean change scores at three and six months. The mean change for both instruments was well above their MCIDs. These results have policy implications in that they demonstrate the validity and responsiveness of both the SF-6D and WOMAC derived HUI3 following knee arthroplasty and thus can be used to measure HRQL for incorporation into cost-utility analyses to inform policy and treatment decisions. However, in this sample, the WOMAC derived HUI3 demonstrated a greater mean change and therefore could result in a more favourable cost-utility analyses given the potential to demonstrate a greater incremental net-benefit relative to using the SF-6D to calculate quality-adjusted life years. However, these findings are not uncommon.[90, 91]

The physical component summary score of the SF-36 following knee or hip replacement surgery has been shown previously to have similar standardized response means when compared with hip and knee-specific instruments and decreased responder burden allowing for the

recommendation of its use as a joint-specific measurement scale.[128] Bombardier et al. hypothesised that both the WOMAC and the SF-36 measure distinct but complementary aspects health following knee replacement surgery and found that the WOMAC discriminated better between subjects with different degrees of knee disability following treatment while the SF-36 discriminated better between subjects with varying levels of self-reported health status and co-morbidity.[129] The use of both a disease specific and general health measure has also been shown to be useful in other musculoskeletal conditions specifically in the assessment of shoulder function. Given that the HUI3 was derived from the WOMAC domain scores, this could lend some explanation as to the reason why it was better able to discriminate HRQL between patients improved and those that did not relative to the SF-6D.[130]

These results are limited given the HUI3 itself may not have the same discriminative properties as the WOMAC derived HUI3. The WOMAC is a disease-specific instrument and therefore may have inflated the HUI3 utility or conversely, although the regression model used in the derivation of the HUI3 is strictly a statistical relationship, the WOMAC does not include all of the HRQL domains encompassed within the HUI3 such as vision, hearing, speech and cognition and therefore may potentially be missing the scope covered within the utility score. The determination of whether or not there is a difference between the WOMAC derived HUI3 and the HUI3 itself is beyond the scope of this paper.

Another limitation of the study was the absence of transition questions for use as the external criteria in the dichotomous response models. The 20% WOMAC pain score cut-point has been evaluated and validated but work is still in progress in an attempt to determine a clinically meaningful cut-off for the OKS. As a result, a range of cut-points was used to identify the most appropriate measure. In addition, because this was a secondary data source, the type of

knee replacement performed (i.e. UKA vs. TKA) was not available for every patient involved in the study; however, this is not expected to have any effect on these analyses.

The WOMAC derived HUI3 and the SF-6D appear to be valid measures of HRQL in this sample. The HUI3 was highly responsive to change and able to clearly discriminate between both health states. The SF-6D was not as responsive but demonstrated a good ability to identify patients who had improved. We therefore would encourage the use of the WOMAC derived HUI3 utility index either on its own or in conjunction with the global SF-6D health utility measure for use in evaluating HRQL following knee arthroplasty for the treatment of osteoarthritis.

Figure 2-1. Inclusion and exclusion of patients

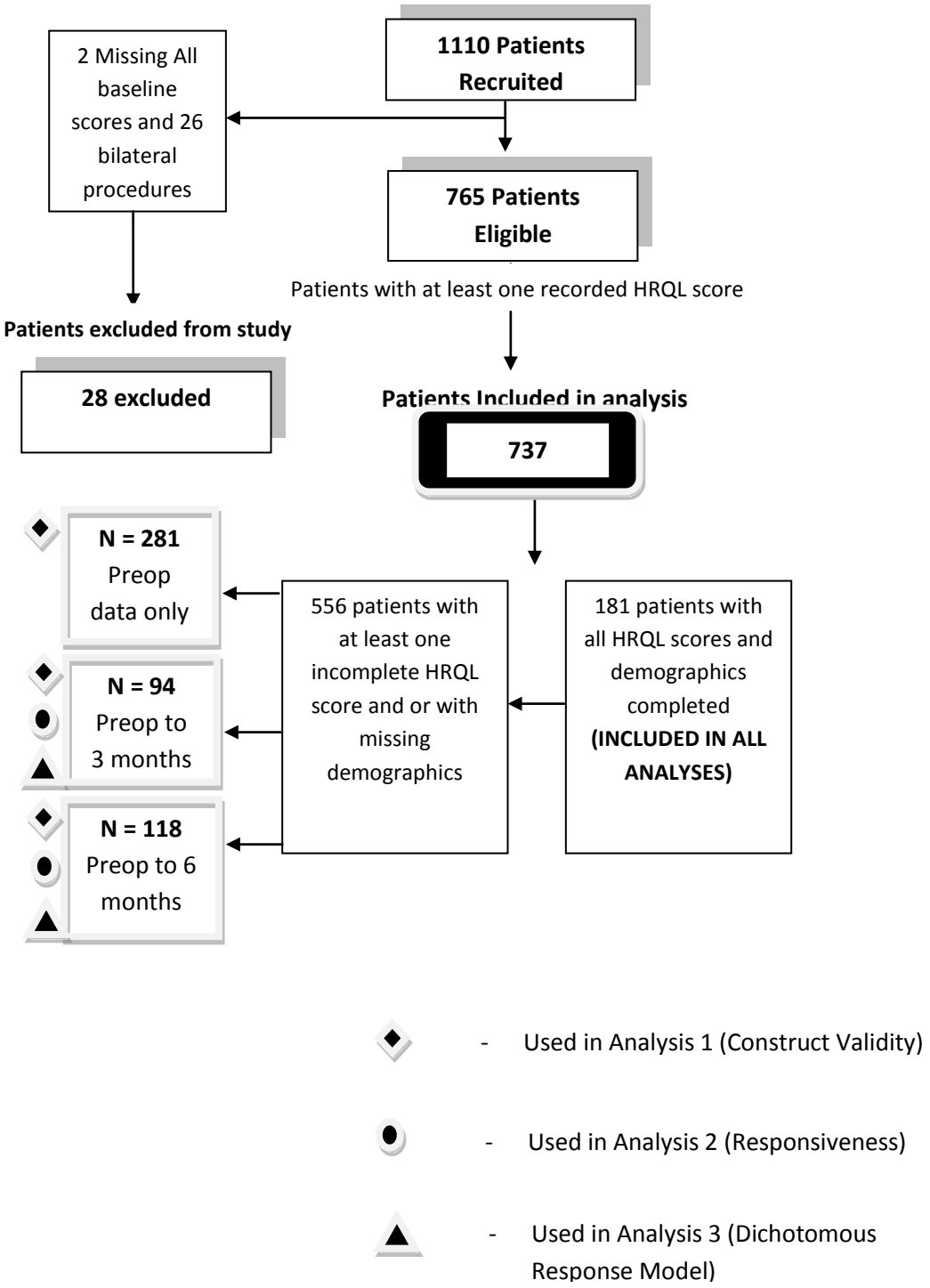


Table 2-1. Comparison of baseline HRQL between those with complete responses and those with incomplete information

	Complete Information* Mean (SD) (N = 181)	Incomplete Information* Mean (SD) (N = 556)	P-value
WOMAC			
Total	53.62 (16.89)	55.73 (15.75)	0.15
Pain	11.33 (3.47)	11.76 (3.38)	0.35
Stiffness	4.75 (1.75)	4.78 (1.67)	0.55
Function	37.54 (12.57)	32.15 (16.16)	,0.001ǂ
OKS	40.95 (7.98)	41.86 (7.60)	0.48
SF-6D	0.62 (0.12)	0.59 (0.11)	0.02ǂ
HUI3	0.48 (0.17)	0.46 (0.16)	0.90
MCS	49.23 (11.35)	48.15 (11.50)	0.86
PCS	32.75 (7.47)	30.60 (7.49)	0.47

*Incomplete information – data collected for at least one time point but not at all three; i.e. preop, 3 and 6 months

*Complete information – data collected at all three time points

Table 2-2. Pearson's R correlation between utility instruments, OKS, and WOMAC scores

	OKS	WOMAC TOTAL	WOMAC PAIN	WOMAC STIFF	WOMAC FUNC	SF-12 PCS	SF-12 MCS
PREOP							
SF-6D	-0.66	-0.64	-0.59	-0.41	-0.64	0.57	0.63
HUI3	0.62	-0.95	-0.85	-0.53	-0.96	0.36	0.42
3 MO							
SF-6D	-0.66	-0.70	-0.64	-0.44	-0.70	0.74	0.66
HUI3	-0.75	-0.88	-0.82	-0.50	-0.89	0.55	0.46
6 MO							
SF-6D	-0.68	-0.69	-0.66	-0.49	-0.69	0.72	0.69
HUI3	0.66	-0.94	-0.83	-0.67	-0.95	0.58	0.52

Table 2-3. Mean scores and mean change scores for each questionnaire at 3 and 6 months post-operatively

Measure	Preop Mean (SD)	3 mo f/up Mean (SD)	6 mo f/up Mean (SD)	Mean change 0 - 3 months Mean (SD)	Mean change 0 - 6 months Mean(SD)
WOMAC					
Total	55.67 (16.46)	26.01 (18.01)	21.39 (17.15)	-29.14 (21.54)	-33.26 (21.54)
Pain	11.70 (3.52)	5.24 (3.94)	4.17 (3.71)	-6.42 (4.75)	-7.39 (4.64)
Stiffness	4.82 (1.70)	2.87 (1.72)	2.46 (1.61)	-1.93 (2.18)	-2.36 (2.13)
Function	39.15 (12.20)	17.77 (13.28)	14.86 (12.59)	-20.88 (15.73)	-23.38 (15.89)
HUI3	0.46 (0.17)*	0.69 (0.14)ǂ	0.71 (0.12)*	0.22 (0.19)	0.25 (0.18)
SF-6D	0.59 (0.12)*	0.74 (0.14)ǂ	0.75 (0.13)*	0.13 (0.15)	0.15 (0.16)
OXFORD (OKS)	41.79 (7.89)	27.20 (9.47)	23.67 (8.85)	-14.17 (11.51)	-17.98 (11.06)
PCS	31.50 (7.59)	39.68 (9.35)	41.34 (9.41)	7.89 (10.23)	10.01 (10.84)
MCS	47.85 (11.62)	52.90 (9.56)	53.20 (9.49)	4.34 (12.04)	4.64 (12.68)

*P value of < 0.001 between HUI3 and SF-6D mean scores preoperatively and at 6 months post-surgery

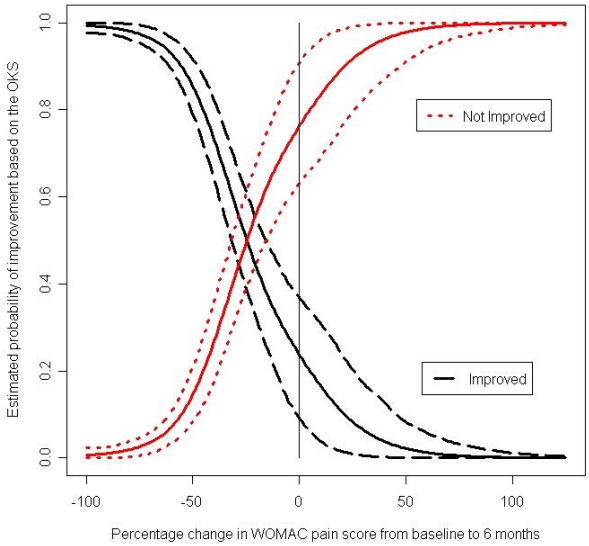
ǂ P value of 0.007 between HUI3 and SF-6D mean scores at the 3 month assessment time post-surgery

Table 2-4. Effect size (ES) and standardized response mean (SRM) at 3 and 6 months post-surgery

	3 Months		6 Months	
MEASURE	**ES (95% CI)**	**SRM (95% CI)**	**ES (95% CI)**	**SRM (95% CI)**
Disease-Specific				
WOMAC	-1.86 (-2.06,-1.67)	-1.36 (-1.54,-1.20)	-2.12 (-2.34,-1.92)	-1.55 (-1.72,-1.40)
PAIN	-1.93 (-2.15,-1.71)	-1.36 (-1.53,-1.18)	-2.25 (-2.49,-2.03)	-1.60 (-1.74,-1.45)
STIFFNESS	-1.15 (-1.30,-0.99)	-0.89 (-1.02,-0.78)	-1.49 (-1.66,-1.32)	-1.11 (-1.24,-0.99)
FUNCTION	-1.80 (-1.99,-1.59)	-1.33 (-1.49,-1.17)	-2.01 (-2.21,-1.82)	-1.48 (-1.66,-1.32)
OXFORD	-1.86 (-2.07,-1.66)	-1.24 (-1.38,-1.10)	-2.41 (-2.63,-2.18)	-1.63 (-1.81,-1.48)
Generic				
HUI3	1.34 (1.17,1.52)	1.14 (0.97,1.33)	1.52 (1.35,1.68)	1.36 (1.21,1.52)
SF-6D	1.12 (0.95,1.30)	0.84 (0.71,0.99)	1.33 (1.16,1.51)	0.96 (0.82,1.12)
PCS	1.08 (0.92,1.23)	0.77 (0.67,0.89)	1.43 (1.25,1.63)	0.93 (0.81,1.06)
MCS	0.38 (0.27,0.49)	0.36 (0.26,0.47)	0.42 (0.30,0.54)	0.37 (0.26,0.47)

Figure 2-2. Results of the dichotomous response model of the association between percent change in WOMAC pain scores and the OKS. The solid lines represent the fitted model and the dotted lines represent the 95% confidence intervals

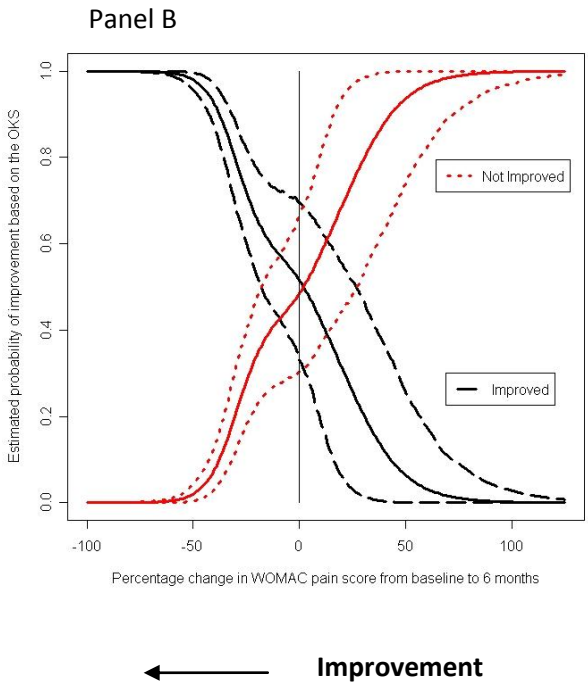
Panel A



Improvement

Panel A - Patients whose scores improved on the OKS by at least 20% were classified as 'improved', while those whose scores did not improve by at least 20% were defined as 'not improved'.

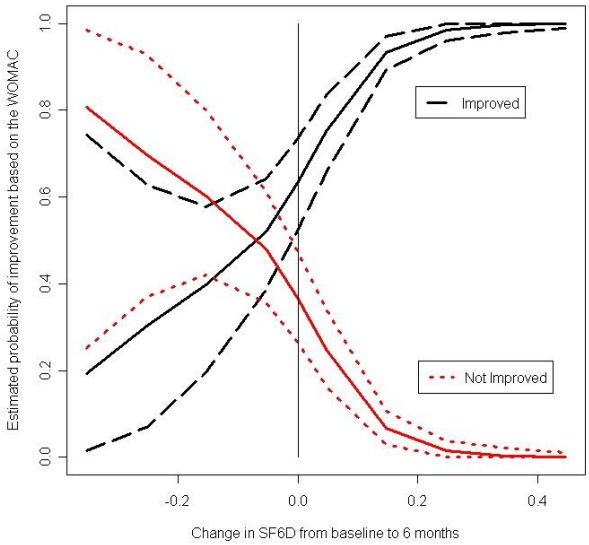
Figure 2-2. Results of the dichotomous response model of the association between percent change in WOMAC pain scores and the OKS. The solid lines represent the fitted model and the dotted lines represent the 95% confidence intervals



Panel B - Patients whose scores improved on the OKS by at least 10% were classified as 'improved', while those whose scores did not improve by at least 10% were defined as 'not improved'.

Figure 2-3. Results of the dichtomous response model of the association between a change in SF-6D scores with both the OKS and WOMAC. The lines represent the fitted model and the dotted lines represent the 95% confidence intervals

Panel A

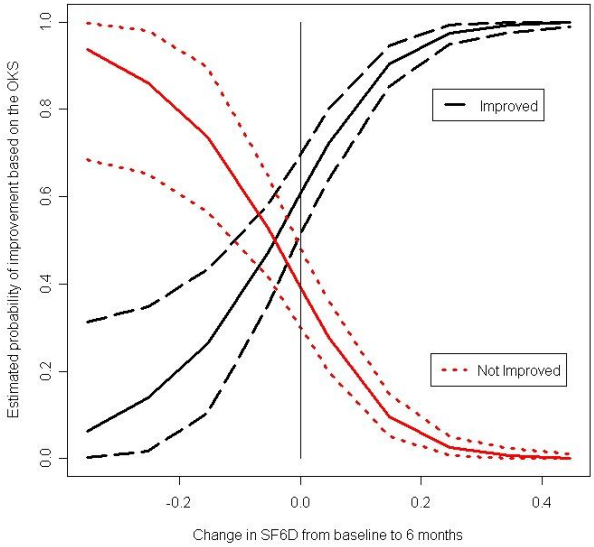


Improvement →

Panel A - Patients whose scores improved by at least 20% on the WOMAC pain score were classified as 'improved', while those whose scores did not improve by at least 20% were defined as 'not improved'.

Figure 2-3. Results of the dichotomous response model of the association between a change in SF-6D scores with both the OKS and WOMAC. The lines represent the fitted model and the dotted lines represent the 95% confidence intervals

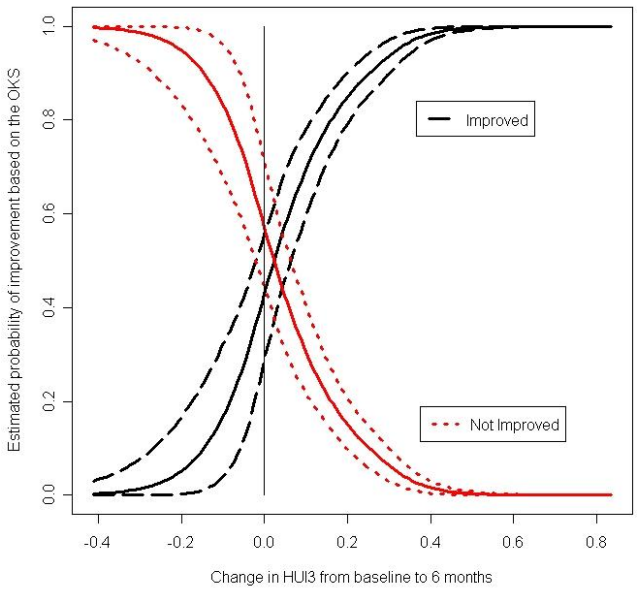
Panel B



Improvement →

Panel B - Patients whose scores improved on the OKS by at least 20% were classified as 'improved', while those whose scores did not improve by at least 20% were defined as 'not improved'.

Figure 2-4. Results of the dichotomous response model of the association between a change in HUI3 score with the OKS. The lines represent the fitted model and the dotted lines represent the 95% confidence intervals



Improvement →

Figure 2.4 - Patients whose scores improved by at least 20% on the OKS were classified as 'improved', while those whose scores did not improve by at least 20% were defined as 'not improved'.

CHAPTER 3

COMPARISON OF HRQL BETWEEN UNICOMPARTMENTAL KNEE ARTHROPLASTY AND TOTAL KNEE ARTHROPLASTY FOR THE TREATMENT OF OA

3.1 FORWARD

This chapter will be submitted for publication, under the same title, in the journal of Orthopaedics and Related Research. The candidate is first author of this manuscript which is co-authored by Maja Grubisic, a statistician who provided statistical consultation and SAS programming in the analysis of the data. The other co-authors included Dr. Richard Kendall, an orthopaedic surgeon who provided surgical expertise and clinical knowledge, Dr. Larry Lynd, the candidate's supervisor who helped in the development of the conceptual framework and methodology of the study, and Drs. Carlo Marra and Linda Li, members of the supervisory committee who also provided comments regarding the manuscript during its preparation.

The candidate was responsible for the statistical analysis, the development of the hypothesis, and wrote the final manuscript.

3.2 INTRODUCTION

Within the last thirty years, advances in technique and design have brought unicompartmental knee replacement (UKA) into the mix as a viable and successful treatment for arthritis affecting only one compartment of the knee.[47, 57, 69] However, some surgeons remain

sceptical towards the use of UKA due to the lack of head-to-head comparisons with TKA.[131] A randomized control trial done by Newman et al. is to our knowledge the only well-controlled trial done directly comparing UKA with TKA.[81] Others have compared different preoperative baseline characteristics and varying degrees of disease progression within the two groups making it difficult to properly identify the true, if any, outcome differences between UKA and TKA.[71, 83] Furthermore, primary outcomes were predominantly classified as success or failures defined by need for revision or cost-effectiveness while others use time to recovery, patient perceptions following surgery or one dimensional aspects of knee function.[70, 72-80] Binary and one dimensional outcomes such as these tend to exclude relevant data making patient self-assessed HRQL questionnaires, which are used with increasing frequency for the evaluation of the effectiveness of treatment, potentially more reflective of the overall success of the procedure at least from the patient's point of view.[32] The objective of this study was therefore to compare the effectiveness of the Biomet Oxford® unicompartmental implant (Warsaw, Indiana) to the Wright Advance® Medical total implant (Memphis, Tennessee) over the first 6 months post-implantation. For this study, effectiveness was evaluated using two condition-specific questionnaires: the Western Ontario and McMaster Osteoarthritis index (WOMAC) and the Oxford Knee Score (OKS).

3.3 PATIENTS AND METHODS

A retrospective analysis was conducted to evaluate outcomes following knee arthroplasty on data collected between May 2002 and February 2007 on all patients requiring either unicompartmental or total knee arthroplasty, as determined at the time of consultation by the orthopaedic surgeon consulted. All 1110 procedures were performed at Richmond General

Hospital in Richmond, British Columbia, Canada by one of five surgeons. Once a decision to perform surgery was made and consent was obtained the patient completed the WOMAC, OKS and SF-12 preoperatively and then subsequently at 3 and 6 months post-surgery. A second data set was used in conjunction with these data consisting of 258 patients who underwent unicompartmental knee arthroplasty all of which were performed by one surgeon at Richmond General Hospital from July 2001 to December 2008. Patients completed the WOMAC, OKS and SF-12 preoperatively and at 3 and 6 months post-operatively.

The WOMAC is a 24-item osteoarthritis specific instrument that was originally designed for use in clinical trials in patients with OA of the knee or hip.[107] It is comprised of questions designed to measure quality of life on three arthritis-specific domains: five questions specific to pain, two questions specific to stiffness and seventeen questions specific to physical function. The Likert 3.1 version was used in this study with the score for each individual domain being calculated by summing the numbered responses corresponding to the domain specific questions and a total score calculated by summing all three domain-specific scores. The score for each item ranged from 0 (full health) to 4 (extreme disability).[113]

The OKS is a 12-item questionnaire in which a single score ranging from 12 (full health to 60 (extreme disability) is derived from patient's self-reported responses to questions relating to pain, knee function and day to day tasks and activities. It has been validated and recommended for use in the evaluation of health related quality of life following knee arthroplasty.[86, 110]

Lastly, the SF-12 is a generic quality of life instrument designed to measure the general health of the target population. It consists of 12 questions encompassing seven domains: general health, physical function, role limitations, pain, mental health, vitality and social functioning.[111]

It is comprised of both a mental component score (MCS) and a physical component score (PCS) that range from 0 (extreme disability) and 100 (full health).

Patients with OA affecting only one compartment of the knee were eligible for UKA in which the Oxford phase III prosthesis was used. It offers a fully congruent mobile bearing that is completely unconstrained and therefore in theory allows for an evenly distributed allocation of weight throughout the joint.[53] The indications used for UKA were: single compartment OA, no inflammatory disease, normal ligaments, and a flexion deformity of less than 15°. The Wright Advance Medical implant utilizing a medial pivot bearing was used for all TKAs.[132, 133] Procedures were performed either under regional or general anaesthetic and tourniquets were typically used. The Oxford implant was implanted through an MIS type of approach while TKA was typically done either through a traditional medial parapatellar approach, intravastus or subvastus approach. Femoral and tibial components were both cemented with Simplex-P with tobramycin (Stryker Kalamzoo Mi) cement. Patellar resurfacing with TKA was at the surgeon's discretion. All patients were mobilized with physiotherapy on the operative day (day 0) with discharge typically between one to three days post-operatively. Upon discharge outpatient physiotherapy was instituted for all patients. Postoperative care and rehabilitation were identical in both groups and follow-ups were performed by the surgeon who performed the knee replacement. Consent for all patients was obtained upon referral for knee replacement and before entry into the study.

3.3.1 Statistical analysis

Demographic variables and quality of life between the two groups were compared using independent samples *t*- test for continuous variables and Pearson's Chi-Square test for categorical variables. Since the measurements that were made at several time points (baseline, 3

months and 6 months) for each patient were interdependent, standard linear modeling cannot be used, as it assumes the errors of each observation are independent with a constant variance. Conversely, hierarchical linear models (HLM) estimate the errors for each patient separately, therefore to investigate the mean longitudinal change, the data were analyzed using a three level hierarchical linear model. In this three level model, the base layer represents time (i.e. baseline, 3 and 6 months follow-up), which is nested within patients who represent level two. Patients are then clustered with surgeon, which represents the third level of the hierarchical model. We used this method to model the total WOMAC scores and OKS as a function of group (either TKA or UKA) and time. The model intercept was treated as a fixed effect because it was assumed that surgeons would perform similarly given that they all have the capability to execute both types of procedures, all of which were done at the same institution under identical protocols, using the same indications and instrumentation. Slope was treated as a random effect and the time effect was assumed to be linear. To investigate how differences between treatments changes over time, an interaction of time with group was incorporated into the model. All variables relating to outcomes following knee arthroplasty were included. The model was adjusted for age, sex and prosthesis type, which was the last of the covariates to be added. For patients with bilateral knee replacements, only the first procedure was evaluated to ensure that the results reflected the outcome of the index surgery. All analyses were performed using SAS 9.1 statistical software (SAS Institute; Cary, NC). Statistical significant was defined as $p < 0.05$.

3.4 RESULTS

Seven hundred and forty-two patients (67%) of the 1110 UKA or TKA's performed were included in the analysis of which 316 (43%) were UKAs and 426 (57%) TKAs. A total of 368

patients were excluded, two (0.5%) were revision surgeries, two (0.5%) of patients were missing all baseline HRQL scores, 26 (7%) of patients had bilateral surgeries in which case only the first procedure was evaluated. The remaining 338 (92%) were missing demographic or surgery related data such as surgeon, age, sex or type of procedure. Patients undergoing TKA were significantly older ($p = 0.02$) and both UKA and TKA groups were more likely to be female ($p = 0.01$ and $p < 0.001$, respectively); however, there were no significant differences in pre-operative HRQL between the two groups (Table 3-1).

Table 3-2 displays the parameter estimates and 95% confidence intervals of the model. Patients showed significant improvement six months following knee arthroplasty with an average change of 6.2 out of a total of 96 points and 3.3 out of a total of 60 points on the WOMAC and OKS, respectively. Age was not significant in the model and males improved significantly more compared to females on the total WOMAC, i.e. -3.6 out of a total of 96 points and the OKS, i.e. -2.0 out of a total of 60 points, following surgery. Prosthesis type (TKA/UKA) was not significant in the model ($p = 0.5$ and 0.8 for both the total WOMAC and OKS, respectively) and therefore patients who underwent TKA had similar HRQL at 6 months follow-up compared to patients who underwent UKA. There was also no significant difference in treatment effect overtime between the two groups as seen through the time and treatment interaction term for both the total WOMAC and OKS ($p = 0.2$ and $p = 0.1$, respectively)

The estimated adjusted mean differences of the total WOMAC and OKS between the UKA and TKA groups at each time point are shown in table 3-3. Although the differences between the two groups were not significant; at baseline the UKA group reported on average 1.8 points better on the total WOMAC and 1.1 points better on the OKS compared to the TKA group. At 3 months the UKA group had improved by 0.1 points more on the total WOMAC and

0.2 points on the OKS compared to patients that underwent TKA and at 6 months the TKA group had improved by 1.5 points more on the total WOMAC and 0.7 points on the OKS compared to the UKA group.

3.5 DISCUSSION

Many surgeons remain sceptical about the implementation of UKA as a treatment for osteoarthritis affecting one compartment of the knee due to its history of the increased need for revision commonly caused by luxation of the prosthesis, misalignment of the joint translating into progression of the disease into the untreated part of the knee as well as the technically advanced nature of the procedure.[64, 134] This is the first study to our knowledge to compare outcomes following UKA and TKA using disease-specific patient self-reported questionnaires with adequate sample size and comparability at baseline between both groups. Using a hierarchical linear model, we were able to account for the correlation within patients over time and the variation within surgeons while adjusting for significant covariates to show that UKA does not differ from TKA with respect to pain, function, stiffness and daily knee-specific activities following surgery. Males had greater improvement in HRQL as compared to females and age was not a significant factor in this sample. On average patients showed significant improvement at 3 and 6 months following both UKA and TKA and there was no significant difference in outcomes between the two groups with respect to the WOMAC and OKS.

Previous studies have shown equivocal results in outcomes related to UKA and TKA with the exception of one study done by Newman et al.[81] In this study, 102 patients, all of which were suitable for UKA, were randomized to receive either a UKA with the St. Georg Sled

(Waldemar Lik, Hamburgh, Germany; n=50) or a TKA with the Kinematic (Howmedica, Rutherford, New Jersey) prosthesis (n=52). With knee flexion and pain relief measured using the Bristol Knee Score (BKS) as one of the primary outcomes, 68% of patients randomized to receive a UKA had achieved normal flexion (≥120 degrees) after five years and only 17.8% in the TKA group suggesting that UKA is superior to TKA. Further, after five years the UKA group had a mean BKS of 91.1 which was indicative of better function than a mean of 86.7 in the TKA group. Patients were subsequently reviewed at a minimum of 15 years following surgery with 78% of UKA patients had normal flexion and 38% in the TKA group.[82] In the UKA group, 71.4 % of patients reported excellent results on the BKS whereas only 52.6% reported the same in the TKA group.

However, the Bristol knee score may not be reflective of the overall improvement in quality of life of the patients and the latter offers only one dimension of relief and amelioration bringing to light the importance of proper outcome measures to adequately reflect these changes. Health related quality of life questionnaires encompass individual's responses relevant to physical, mental, emotional and daily activities of life and enables clinicians to include the broader implications of the disease. It is therefore evident that the assessment of outcomes following knee arthroplasty should focus more on the overall health and quality of life from the patient's perspective.[135]

A recent study performed by Walton et al. attempted to account for patient-perceived change in outcome after either UKA or TKA using the Oxford Knee Score questionnaire (OKS) which assesses function and daily activities, the modified Gimby score which assesses levels of patient physical activity as the primary outcomes following surgery along with patient's levels of sports participation and working status.[83] The UKA group performed better on the OKS and

modified Gimby score with mean scores at 1 year of 22.2 and 3.9, respectively, compared to 24.5 and 2.8, respectively, in the TKA group. UKA patients were significantly more likely to increase or maintain their pre-operative level of sporting activity following surgery compared to TKA patients ($P < 0.001$). However, the authors do not report preoperative OKS and modified Gimby scores for both groups and therefore it is unsure if they are clinically or statistically significantly similar at baseline making it difficult to draw any concrete conclusions from the comparison. For this reason a comparison study properly evaluating patient-centered HRQL outcomes between UKA and TKA is needed.

Similarly, Anish et al. conducted a five year follow-up study comparing fifty four patients who underwent UKA with fifty four patients who underwent a TKA. All patients selected for a UKA had anteromedial OA treated with the Oxford Biomet® implant.[51] Unicompartmental knee arthroplasty patients were then matched by age, BMI, preoperative knee score, function score, degree of active range of motion (ROM), and gender to unilateral TKA's that were identified retrospectively by one of the authors. Unicompartmental OA was not a requirement for the entry of TKA patients into the study. The five year survivorship rate based on revision for any reason was 88% for UKA and one hundred percent for TKA, but the mean postoperative ROM was significantly better in the UKA group compared to the TKA group (99° vs 104°; $p = 0.001$).[71] Given the uncertainty surrounding the criteria used and the subjectivity inherent in the matching of the two groups by the author, it is unclear how comparable the UKA and TKA groups were at baseline in terms of function and quality of life, as these data are not provided.

The present study was a retrospective analysis using data derived from a secondary source and therefore randomisation could not be performed. As a result, not all patients were

eligible for a UKA yet, we believe the similarity in HRQL at baseline between the UKA and TKA groups at least partly accounts for this shortfall. Another limitation is the lack of long term follow-up in this sample. Given that the most notable change in HRQL generally occurs within six months following knee arthroplasty, we are confident that the six month follow-up allows for an adequate amount of time to be representative of the final outcome of the patients.[42, 93]

Unicompartmental knee arthroplasty appears to be comparable to total knee arthroplasty with respect to HRQL measured by two disease-specific questionnaires in the total WOMAC and OKS. This study therefore lends support for the use of UKA in patients with the appropriate indications.

Table3-1. Baseline characteristics of unicompartmental and total knee arthroplasty patients

Variables	Uni Knee N = 316	Total Knee N = 426	P-Value
Sex	179 (F)	279(F)	p = 0.01 and p < 0.001, respectively
	Mean (SD)	Mean (SD)	
Age	69.0 (9.5)	70.4 (9.0)	p = 0.03
Total WOMAC	54.2 (15.3)	56.2 (16.6)	p = 0.2
Pain	11.5 (3.3)	11.8 (3.5)	p = 0.2
Stiffness	4.8 (1.6)	5.0 (1.7)	p = 0.2
Function	37.7 (11.8)	39.5 (12.3)	p = 0.4
Oxford Knee Score	41.3 (7.3)	42.3 (7.9)	p = 0.07
PCS	32.7 (7.3)	31.7 (7.3)	p = 0.4
MCS	48.8 (11.8)	47.5 (11.3)	p = 0.6

Table 3-2. Estimated parameters and 95% CIs of the mixed effects model comparing HRQL between UKA and TKA

Variables	Total WOMAC	OKS
Intercept	60.8* (53.0, 68.7)	43.2* (39.4, 47.0)
Group (UKA)	-1.8 (-4.2, 0.7)	-1.1 (-1.9, 0.8)
Time	-6.2* (-7.0, -5.4)	-3.3* (-3.8, -2.8)
Time x (UKA)	0.5 (-0.2, 1.3)	0.1 (-0.09, 0.7)
Sex (Male)	-3.6* (-5.6, -1.5)	-2.0* (-3.0, -1.0)
Age	-0.1 (-0.2, 0.03)	-0.01 (-0.07, 0.04)

*Significance of $p < 0.001$

Table3-3. Differences in HRQL scores between unicompartmental and total knee arthroplasty patients at each assessment

Scale	Estimates of Differences at Baseline (95% CI)	Estimates of Differences at 3 mo (95% CI)	Estimates of Differences at 6 mo (95% CI)
WOMAC total	-1.8 (-4.2, 0.7)	-0.12 (-2.5, 2.2)	1.5 (-2.5, 5.5)
OKS	-1.1 (-2.3, 0.06)	-0.2 (-1.4, 1.0)	0.7 (-1.3, 2.8)

CHAPTER 4

THE EFFECT OF SOCIOECONOMIC STATUS ON HRQL OUTCOMES FOLLOWING UNICOMPARTMENTAL AND TOTAL KNEE ARTHROPLASTY

4.1 FORWARD

This chapter will be submitted for publication, under the same title, to the Journal of Athroplasty. The candidate is first author of this manuscript which is co-authored by Maja Grubisic, a statistician who provided statistical expertise and programming knowledge in the analysis of the data. Dr. Richard Kendall an orthopaedic surgeon who provided surgical input, Dr. Larry Lynd, the candidate's supervisor who proposed the use of the HLM model, and Drs. Carlo Marra and Linda Li, members of the supervisory committee.

The candidate's role in this manuscript was the general direction of the work, the study hypothesis, data preparation, the statistical analysis, and the writing of the final manuscript.

4.2 INTRODUCTION

Osteoarthritis (OA) of the knee is the most common form of arthritis causing debilitation, pain and lack of mobility in patients suffering from the disease.[1, 2] Knee arthroplasty has been shown to be an effective treatment for symptomatic OA of the knee[30, 40, 42, 47, 77, 136, 137] yet controversy exists over the use and of unicompartmental versus total knee arthroplasty. In addition to preoperative WOMAC pain and function scores, several other factors have been associated with HRQL following TKA, including BMI, age, other co-morbid conditions, mental

health status and socioeconomic status. However, few explanatory variables have been identified that may predict health related quality of life (HRQL) outcomes specifically patient sub-groups who may respond more favourably to UKA as opposed to TKA. Lower socioeconomic groups have been shown to have worse outcomes following other surgical interventions specifically after liver and renal transplantation yet little work has been done on the effects of socioeconomic status (SES) on outcomes following knee

arthroplasty.[97, 98] A study performed by Davis et al. showed that lower income patients had a greater need for TKA in terms of poorer knee function and pain prior to surgery but these patients were able to compensate and obtain similar outcomes following surgery.[99] To date no studies have evaluated the effects of SES on patients undergoing UKA or on the effects of SES between patients undergoing UKA or TKA. The purpose of this study was to evaluate differences between UKA and TKA with respect to the effects of SES on HRQL pre-and post-operatively.

4.3 PATIENTS AND METHODS

A retrospective analysis to evaluate the effects of SES on HRQL was conducted on data collected on all consenting patients undergoing either a UKA or TKA between July 2000 to November 2008 at Richmond General Hospital. The Western Ontario and McMaster Osteoarthritis index (WOMAC) and Oxford Knee Score (OKS), which were used as the primary measures of HRQL, in conjunction with the SF-12, were completed preoperatively, and at 3 and 6 months following surgery. All UKAs were implanted with the Biomet® Oxford implant (Warsaw, Indiana) and the Advance® Wright Medical implant (Memphis, Tennessee) was used

for all TKAs at the same institution by five different surgeons. Ethics approval was received from the University of British Columbia Research Ethics Board (#C04-0577) and research approval was received from the Richmond Health Services Research Advisory Committee.

Patients with OA affecting only one compartment of the knee were eligible for UKA in which the Oxford phase III prosthesis was used. It offers a fully congruent mobile bearing that is completely unconstrained and therefore in theory allows for an evenly distributed allocation of weight throughout the joint.[14] The indications used for UKA were: single compartment OA, no inflammatory disease, normal ligaments, and a flexion deformity of less than 15°. The Wright Advance Medical implant utilizing a medial pivot bearing was used for all TKAs.[16, 138] Procedures were performed either under regional or general anaesthetic and tourniquets were typically used. The Oxford implant was implanted through an MIS type of approach while TKA was typically done either through a traditional medial parapatellar approach, intravastus or subvastus approach. Femoral and tibial components were both cemented with Simplex-P with tobramycin (Stryker Kalamzoo Mi) cement. Patellar resurfacing with TKA was at the surgeon's discretion. All patients were mobilized with physiotherapy on the operative day (day 0) with discharge typically between day 1 and 3 post-operatively. Upon discharge outpatient physiotherapy was instituted for all patients. Postoperative care and rehabilitation were identical in both groups and follow-ups were performed by the surgeon who performed the knee replacement. Consent for all patients was obtained upon referral for knee replacement and prior to entry into the study.

4.3.1 Outcome measures

The WOMAC and OKS were used as the primary measures in evaluating the effects of SES on patient's pain and function following knee arthroplasty. The WOMAC is a 24-item osteoarthritis specific instrument that was originally designed for use in clinical trials in patients with OA of the knee or hip.[107] It has been vigorously tested for validity, reliability and responsiveness and is comprised of three domains; pain (5 questions), stiffness (2 questions) and physical function (17 questions). The Likert 3.1 version was used in this study with each domain score calculated by summing the coded responses corresponding to the domain specific questions and a total score calculated by summing the scores for each domain. The scores for each item ranged from 0 (full health) to 4 (extreme disability).[113]

The OKS is a 12-item questionnaire that has been previously validated and is recommended for the evaluation of HRQL following knee arthroplasty.[86] A single score is derived from patient's self-reported responses to questions relating to pain, knee function and day to day tasks and activities.[110]

To assess the overall health of these patients the SF-12, which is a generic multi-disease state tool that measures the general health of the target population, was used. It consists of 12 questions encompassing seven domains: general health, physical function, role limitations, pain, mental health, vitality and social functioning.[111] It is comprised of both a mental component score (MCS) and a physical component score (PCS) that range from 0 (worst health state) and 100 (perfect health).

4.3.2 Socioeconomic status

Individual measures of SES were based on the median estimates of annual neighbourhood household incomes. This was done using Statistics Canada Postal Code Conversion File, which determines the dissemination area for each patient's current residency. The neighbourhood median annual income could then be derived from the province of British Columbia's census data. To evaluate whether patients in lower socioeconomic groups differ with respect to disease-specific and generic HRQL prior to knee arthroplasty and to identify if these patients have poorer post-operative outcomes, income was categorized into four groups. The first income group consisted of patients with an annual median household income of < $40 000; the second was defined by an annual median household income between $40 000 and $55 000; the third consisted of patients with an annual median household income of greater than $55 000 and less than $70 000 and the fourth group consisted of an annual median household income of greater than $70 000. One-way ANOVA was used to compare the mean HRQL scores between socioeconomic groups. Demographic variables and quality of life between the two treatment groups were compared using independent samples *t*- test and Pearson's Chi-Square test.

4.3.3 Statistical analysis

A hierarchical linear model (HLM), which accounts for the interdependency of each observation, was used to investigate the mean longitudinal changes from the preoperative assessment to 3 and 6 months follow-up.[139] The data were analyzed using a three level hierarchical linear model. The base layer represents time (i.e. baseline, 3 and 6 months follow-up), which is nested within patients who represent level two, and surgeons, each with their cluster of patients, define level three. This method was used to model the total WOMAC scores

and OKS as a function of group (either TKA or UKA) and time. The model intercept was treated as a fixed effect given that each surgeon was assumed to have like abilities in performing surgery, all of which were done at the same hospital, with similar protocols and instrumentation. Slope was treated as a random effect, the time effect was assumed to be linear and income was entered as a categorical variable with group four as the reference group (median annual household income of > $70 000). To identify any differences in outcome between treatment groups with respect to SES, an interaction of group and income was used, with UKA as the reference group. Differences in means were computed for both groups at baseline 3, and 6 months. The model was adjusted for age, sex, prosthesis type and income, which was the last of the covariates to be added. For patients with bilateral knee replacements, only the first procedure was evaluated to ensure that the results reflected the outcome of the index surgery. All analyses were performed using statistical software (SAS 9.1: SAS Institute; Cary, NC). Statistical significance was defined as $p < 0.05$.

4.4 RESULTS

A total of 965 patients were recruited into the original study, two of which reported no HRQL measurements at any of the assessment times and 465 patients were excluded due to incomplete demographic or surgery related data such as surgeon, age, sex or type of procedure. Five hundred patients were included, twelve of which did not report their postal code and therefore were also excluded. This left 488 patients for the final analysis, 152 of which had only preoperative scores, 48 of which reported HRQL measurements at the preoperative assessment and 3 months post-operatively, 91 of which were evaluated preoperatively and at 6 months, and

197 of which reported outcomes preoperatively and at 3 and 6 months using patient-reported HRQL instruments (WOMAC, OKS or SF-12).

The TKA group was significantly older than the UKA group, with means of 70 and 67, respectively and more likely to be female ($p < 0.001$) (Table 4-1). However, there were no significant differences between the two groups with respect to preoperative total WOMAC and OKS scores. To ensure that entering the intercept as a fixed effect was a valid assumption; we ran the model treating the intercept both as a fixed effect and as a random effect. The output did not change in any significant or noticeable way.

To evaluate differences between socioeconomic groups at each assessment time, patients undergoing either UKA or TKA were grouped together (Table 4-2). Table 4-2 shows the categorization of income, which was divided into: group 1, consisting of patients with a median household income of < $40 000; group 2, including patients with a median household income of between $40 000 and $55 000; group 3, comprised patients with a median household income of greater than 55 00 and less than $70 000; and group 4, which was used as the reference group, consisting of patients with a median household income of greater than $70 000. When the unadjusted raw scores were compared at the preoperative assessment the four income groups were significantly different in the total WOMAC scores but not with the OKS ($p = 0.005$ and $p = 0.14$, respectively). However, no significant differences were seen for both the total WOMAC and OKS at three months ($p = 0.2$ and $p = 0.2$, respectively) and six months ($p = 0.5$ and $p = 0.3$, respectively) follow-up.

The estimated parameters from the model and their 95% CIs are displayed in table 4-3. Time and sex had significant effects on both the total WOMAC and OKS. Patients with a median household income of less than $40 000 (group 1) had significantly higher total WOMAC

and OKS scores (larger scores mean increased disability) than the reference group ($p = 0.006$ and $p = 0.01$) at six months postoperatively. On average, group 1 scored 5.1 out of a possible 96 points higher on the total WOMAC and 2.4 out of 60 possible points on the OKS. Income groups 2 and 3 did not differ significantly from the reference group for both the total WOMAC and OKS. In addition, as seen in the treatment and income interaction terms, there were no significant differences between UKA and TKA across income groups.

4.5 DISCUSSION

Due to the continued scepticism over the use of UKA or TKA, surgeons are in need of patient level characteristics that may identify sub-groups who achieve greater improvement with UKA as opposed to TKA. Currently there are no studies that have evaluated the effects of SES on HRQL outcomes following UKA or on potential differences in outcome between UKA and TKA and this study is therefore the first to do so.

In this sample, patients with a lower income had significantly worse unadjusted total WOMAC scores prior to knee arthroplasty (both UKA and TKA) compared to those with higher incomes. However, these patients were able to compensate for this post-operatively, demonstrating similar outcomes between socioeconomic groups at 3 and 6 months following knee replacement. In the hierarchical model, males performed significantly better on the total WOMAC and OKS and patients with a household annual median income of less than $40 000 performed significantly worse on the total WOMAC and OKS at 6 months follow-up. However, no significant differences in HRQL were seen between patients who underwent a UKA versus those who underwent a TKA across income groups and therefore although SES does affect

disease-specific scores at 6 months post-operatively, this effect does not differ between UKA and TKA.

These results are consistent with a recently published study by Davis et al.[99] who studied socioeconomic status in patients undergoing TKA and found that those with lower income appeared to have worse preoperative disease-specific outcomes yet these patients were able to obtain comparable outcomes following surgery. Similarly, Young et al. demonstrated that patients in lower socioeconomic groups were about twice as likely to require TKA but were less likely to receive knee arthroplasty.[96]

In the United States Lower income has been associated with difficulties in accessing health care resources across geographic areas specifically in relation to the prevalence and severity of osteoarthritis requiring total knee arthroplasty. Other reasons for variations in the rate of knee arthroplasty across geographic areas include: knee anthropometrics, the propensity among cultures to seek medical attention and differences in opinion between patients and surgeons on the effectiveness and expectations following knee arthroplasty.[140] It is also thought that patients with lower income are less likely to have private health care insurance, and may therefore wait longer for surgery or forgo treatment all together.[141] Alternatively, Canada relies on a publicly funded health care system which claims to offer equal access to treatment for all socioeconomic groups yet these barriers still exist, making this a complicated issue involving many factors.[142]

The effect of SES on the outcome of other disease states has been studied. For example, lower SES groups have been shown to have worse outcomes after liver[97] and renal transplantation[98] and it has been associated with decreased lung function[143], an increased risk of

cardiovascular disease[144] and hypertension[145]. Preoperative pain and physical function based on the WOMAC scores have also been shown to be strong predictors of outcome at 6 months and two years.[92] In addition co-morbid conditions, BMI, age and mental health status are also highly correlated with outcomes following knee replacement.[93] In this study, the TKA group was significantly older and more likely to be female. Given that in the model males performed significantly better on both the WOMAC total and OKS following surgery, the higher proportion of females, who generally do not do as well as males following knee arthroplasty, may have attenuated the effect of TKA with respect to the HRQL outcomes post-operatively making it more comparable to UKA. However, given the large patient sample in both groups and the comparability at baseline in terms of HRQL we don't foresee this to have confounded the results in a significant way.

Indicators which are non-patient-specific have also been shown to affect potential improvement in patient's HRQL following treatment such as; surgical expertise and repetition. Robertsson et al. found that centers that performed more than 23 cases per year achieved better results.[146] It has also been demonstrated that because UKA is more difficult to perform than TKA with a smaller margin for error and with the use of an unconstrained bearing, an additional hazard of dislocation is introduced, once again making repetition and familiarity important features.[5, 147]

A limitation to the study was that household income was not directly available and therefore had to be derived from the estimates of median household income for each patient's dissemination area of living using Statistics Canada Postal Code Conversion File. As a result patients were categorized into one of four annual neighbourhood median income groups in order to determine if any differences exist with regards to HRQL. Preoperative HRQL was slightly

worse in the lower income group however this was not clinically significant given that these differences were below the minimum 20% change in WOMAC pain scores and half the standard deviation of the mean change score for the OKS, which are considered clinically meaningful thresholds for both instruments.[125, 126] We also did not have access to the education level or the number of individuals within each household which can both have an impact on annual household income. However, use of income as a proxy for SES has been used in previous studies and we believe offers a valid representation of SES for each patient.[148, 149]

Table4-1. Basic demographcis and HRQL scores between UKA and TKA

All Patients Mean (SD)		UKA	TKA	P-value
Age	69.3 (9.4)	66.9 (10.3)	70.4 (8.8)	P = 0.03
Sex	187 (M) 313 (F)	66 (M) 86 (F)	121 (M) 227 (F)	p < 0.001
Total WOMAC	56.0 (15.7)	54.9 (15.4)	56.4 (15.9)	P = 0.97
OKS	42.2 (7.6)	41.4 (7.0)	42.6 (7.8)	P = 0.1

Table4-2. Mean total WOMAC and OKS scores at each time point by income categories

	Preop	N	3 Mo	N	6 Mo	N
			WOMAC			
Income category						
< $40 000	60.8 (16.1)	91	26.8 (18.6)	40	23.5 (21.3)	51
$40 000 - $55 000	55.8 (15.5)	149	28.8 (18.6)	74	21.5 (15.8)	83
>$55 000 - $70 000	53.3 (15.1)	135	26.4 (18.1)	76	19.3 (15.4)	88
>$70 000	54.8 (15.6)	95	21.7 (15.0)	50	19.6 (15.8)	63
P value	0.005		0.18		0.49	
			OKS			
Income category						
< $40 000	43.8 (7.3)	92	27.3 (9.7)	41	25.5 (11.2)	50
$40 000 - $55 000	41.9 (7.5)	151	28.7 (10.1)	74	22.8 (7.7)	83
>$55 000 - $70 000	41.6 (7.4)	140	26.7 (9.1)	79	23.6 (8.2)	90
>$70 000	41.7 (7.9)	97	25.0 (7.9)	51	22.8 (8.0)	63
P value	0.14		0.17		0.29	

Table4-3. Estimated parameters and 95% CIs after adjustment for covariates

Variables	Total WOMAC	OKS
Intercept	58.0* (48.1, 67.9)	42.0* (37.3, 47.0)
Treatment (UKA)	-3.6 (-9.5, 2.2)	-1.1 (-2.6, 0.4)
Time	-6.0* (-7.0, -5.1)	-3.2* (-3.9, -2.6)
Sex (Male)	-3.2ǂ (-5.7, -0.7)	-1.5ǂ (-2.7, -0.3)
Age	-0.05 (-0.20, 0.08)	-0.01 (-0.07, 0.05)
Income (reference group > $70 000)		
<$40 000	5.1ǂ (0.5, 9.8)	-0.30 (-0.70, 0.11))
$40 000- $55 000	0.6 (-3.5, 4.7)	0.8 (-0.8, 2.5)
>$55 000- $70 000	-0.9 (-5.1, 3.3)	0.4 (-1.2, 2.1)
Time x treatment	0.24 (-0.61, 1.10)	0.09 (-0.35, 0.53)
Income group 1 x Treatment (UKA)	0.7 (-7.4, 8.9)	1.0 (-2.9, 4.9)
Income group 2 x treatment (UKA)	4.3 (-3.0, 11.7)	2.6 (-0.9, 6.1)
Income group 3 x treatment (UKA)	3.7 (-5.1, 9.6)	-0.4 (-3.9, 3.0)

*Significance of $p < 0.001$
ǂ Significance of $p < 0.05$

CHAPTER 5

GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 SUMMARY OF STUDY FINDINGS

Measurement of health related quality of life (HRQL) has become increasingly common in the evaluation of interventions used for the management of chronic diseases such as osteoarthritis (OA). In this study, both the SF-6D and WOMAC derived HUI3 appear to be valid instruments and demonstrate good correlation with the disease-specific HRQL questionnaires. Both utility measures had a high probability of discriminating between patients who had improved versus those who had not based on the external criteria and therefore can be appropriate generic preference-based measures used for the evaluation of knee arthroplasty for the treatment of OA.

The validation of HRQL instruments is of particular importance specifically when comparing these outcomes between UKA and TKA. The results of this study provide evidence for the use of UKA as an effective alternative to TKA in terms of patient centered HRQL outcomes for the treatment of OA. It therefore offers surgeons a viable and less invasive option for patients presenting with OA affecting only one compartment of the knee.

Unicompartmental knee arthroplasty did not differ significantly compared to TKA in terms of HRQL across income groups. However, patients receiving a knee arthroplasty (either UKA or TKA) with a median household annual income of less than $40 000 have significantly

lower mean HRQL scores than patients with a median household annual income above this threshold.

5.1.1 Evaluation of the construct validity and responsiveness of generic utility measures

In the analysis assessing the validity and responsiveness of the WOMAC derived HUI3 and SF-6D, both appear to be valid measures of HRQL. They were strongly correlated to the disease specific questionnaires and given that both are generic utilities and therefore measure the same construct, there was strong agreement between the two. In terms of the responsiveness, the magnitude of change was not as great from baseline to 3 and 6 months following surgery on the SF-6D when compared to the HUI3. This was seen in the absolute change scores which translated into smaller effect size and standardized response means on the SF-6D. This could potentially be a result of differences in the range of the scales between the two utilities; -0.36 (representing the worst possible health state) to 1.00 (perfect health) on the HUI3 and 0.3 (worst possible health state) to 1.00 (perfect health) on the SF-6D.

In the study done by Grootendorst et al.[104] which mapped the WOMAC into the HUI3, only 10% of the absolute differences between the predicted and actual baseline HUI3 utility scores were below its MCID of 0.03. The authors therefore concluded that the model was unable to make reliable predictions on individual subject utility scores. However, at the group level, the absolute difference between the predicted and the actual utility scores was not statistically significant ($p < 0.001$) and was therefore able to properly predict HUI3 scores. This is most likely a result of any large prediction errors of individuals within the group cancelling each other out and given that group-level estimates were used in the present study, we can be confident that

the scores are adequately precise. Furthermore, the change scores at 3 and 6 months post-surgery for both measures were greater than their respective MCIDs and therefore clinically meaningful improvement was seen from both utilities.

The results of the dichotomous response model show that the SF-6D had a higher probability of identifying patients who had improved based on the external criteria whereas the HUI3 had a higher probability of identifying patients who had not improved. For example, a -0.2 point change on the HUI3 was associated with a 95% probability of showing no improvement on the external criteria whereas the same magnitude of deterioration on the SF-6D translated into either a 65% or 80% probability of demonstrating no improvement on the external criteria.

However, given that knee arthroplasty, consisting of both UKA and TKA, is a highly effective treatment for OA, the majority of patients showed some form of improvement following surgery. As a result, a small portion of patients were classified as not improved and therefore it is difficult to draw concrete conclusions about the discriminative ability of both utilities with respect to patients showing no improvement post-knee arthroplasty. The HUI3 and SF-6D had a high probability of identifying patients who had improved based on the external criteria (WOMAC and OKS). Both measures have been validated and recommended for use in the evaluation of the treatment of OA and in the evaluation of knee arthroplasty respectively. It therefore can be assumed to adequately categorize patients as improved or not improved. For example, a 0.2 point change on the SF-6D was associated with approximately a 95% probability of improvement based on the external criteria. Similarly, an improvement of 0.2 on the HUI3 was associated with an 80% probability of improvement based on the external criteria.

Previous studies have shown that condition specific preference based scores discriminate better across all levels of disease state whereas generic utilities are limited to differentiating

between only extreme degrees of the condition. McTaggart-Cowen et al. for example, reported on the validity of generic and condition-specific preference-based instruments in terms of asthma control.[91] The Asthma Quality of Life (AQ-5D), which is a condition-specific preference-based instrument that, in the McTaggart-Cowen study, was derived from the Asthma Control Questionnaire (ACQ) and the Asthma Quality of Life Questionnaire (AQLQ(S)) was compared to the HUI3, SF-6D and EQ-5D generic preference-based instruments. Overall the AQ-5D was better able to discriminate across all levels of asthma control whereas the generic utilities were unable to discriminate between the moderate levels. In this study, the HUI3 was derived from patient's age, sex, WOMAC subscales and their interaction terms. Given that the WOMAC domain scores which measure dimensions of pain, stiffness and disability that are specific to the treatment of OA, there is potential for a higher sensitivity to change resulting in the deflation of the utility score. However, the regression model used to derive the HUI3 is strictly a statistical relationship. The HUI3 is a generic preference based utility and although it was derived from a disease-specific HRQL instrument this should not affect the fact that it is measuring the same dimensions as the HUI3 itself.

5.1.2 Evaluating differences in HRQL between UKA and TKA

In the analysis evaluating differences in HRQL between the two types of knee arthroplasties, both the UKA and TKA groups were statistically similar at baseline with regards to HRQL and therefore were assumed to suffer from similar degrees of disease progression. The TKA group however was significantly older and more likely to be female. Factors relevant to the outcome of surgery were included in the hierarchical linear model (HLM) and the inherent

variability within patients and within surgeons were accounted for. This yielded no significant HRQL outcome differences between UKA and TKA at six months post-surgery.

Most of the present literature consists of long-term survivorship studies which focus predominantly on success or failure of the prosthesis.[72, 73, 80] Those that do incorporate HRQL have used one dimensional aspects of knee pain, function and activity level or simply used patient's perceptions on the overall outcome of the procedure.[70, 74-79] Others have compared patient groups that are significantly different at baseline with regards to knee specific questionnaires which may be an indication of varying degrees of disease progression therefore making it difficult to identify any true outcome differences.[71, 83]

5.1.3 Effects of socioeconomic status on HRQL outcomes following knee arthroplasty

In terms of socioeconomic status on the effects of HRQL following surgery, the analysis demonstrates that those with a median annual household income above $40 000 perform significantly better on both the total WOMAC and OKS than patients with a median annual household income below this threshold. Although, these lower income patients performed significantly worse on the unadjusted mean total WOMAC scores preoperatively compared to the three other higher income groups, they were able to compensate for this postoperatively and all four income groups performed similarly at three and six months follow-up. More importantly, this is the first study to evaluate the effects of SES on outcomes following UKA and that these effects do not differ between patients undergoing UKA or TKA.

Studies have demonstrated that low SES is associated with negative effects and the development of many diseases ranging from decreased lung function to an increased risk of acquiring asthma and hypertension.[143, 145, 150] The present literature has also shown that although patients in lower SES groups are more in need of knee arthroplasty than the mid to high range groups in terms of recording worse disease-specific HRQL scores prior to surgery, they are able to compensate and perform similarly on post-operative scores. The results of our analysis were therefore not entirely consistent with previous findings in that patients with a neighbourhood median household income of less than $40 000 performed significantly worse than those above this threshold on both the total WOMAC and OKS scores.

5.2 UNIQUE CONTRIBUTIONS, IMPACT AND IMPLICATIONS

The results presented in the previous chapters contribute significantly to the current body of literature in terms of validating preference-based generic utility measures for use in the evaluation of knee arthroplasty and through the findings relating to HRQL outcomes following knee arthroplasty. Many studies looking at the effectiveness of both UKA and TKA have used a variety of outcome measures such as prosthesis survival, patient perceptions, and one dimensional aspect's of knee function. This is the first well designed study to properly compare HRQL outcomes between UKA and TKA by ensuring similarity between the two groups at baseline. The UKA group was significantly more likely to be female, which may be due to the fact that unicompartmental knee OA predominantly affects the medial part of the knee and this diagnosis is more pronounced in women. The TKA group was significantly older however; both treatment groups consisted mainly of patients above the age of 65 and therefore this patient sample consisted predominantly of the elderly population. There were no significant

preoperative differences between the two groups and we therefore believe surgeons can feel confident in choosing UKA for patients presenting with OA affecting only one compartment of the knee.

Although UKA is a less invasive procedure with lower post-operative morbidity and decreased hospital stay compared to TKA, surgeons are in need of patient level characteristics that may help identify those who respond more favourably to UKA as opposed to TKA and as a result, UKA is not performed as frequently as it could be. Several factors have been identified which have been shown to affect outcomes for patients undergoing either UKA or TKA yet, nothing has been done on the effects of socioeconomic status on UKA compared to TKA. SES has been shown to have significant effects on outcomes in other disease states and this study was therefore the first to do so in UKA compared to TKA.

The primary purpose of using HRQL instruments in assessing outcomes in knee arthroplasty is not specific to the joint but to the overall impact on health, specifically that relating to pain. Valid and responsive measures of health related quality of life are therefore important both in identifying factors that may affect the outcome of knee arthroplasty and in determining the effectiveness of UKA compared to TKA. However the difficulty with using HRQL instruments is in the range of raw scores. These scores are often highly variable across assessment tools with varying psychometric properties and inconsistent levels of validation, making the results difficult to interpret and compare between studies and treatments.[42] Utilities offer a more universal value that is capable of describing a wide variety of health states which may lead to a more structured and reproducible strategy in evaluating the effectiveness of knee arthroplasty for the treatment of OA. Moreover, when assessing the value of treatment from a payer's perspective, typically a generic preference-based measure of HRQL is required that can

be used to prioritize interventions across different disease states. Therefore, in order to compare the effectiveness of knee arthroplasty for the treatment of OA to other treatment initiatives, these previously published studies, which predominantly use the WOMAC disease-specific questionnaire as a measure of HRQL, would have to be re-done using generic health state utilities as outcome measures. This is unrealistic and would be costly and time-consuming. For this reason the WOMAC has been mapped into the HUI3 utility scores.[104] This study is therefore the first to evaluate this novel health state utility in terms of its construct validity and responsiveness in comparison to the SF-12 derived SF-6D utility. Researchers, policy makers and health care providers can now put this utility derivation into use to help decision makers evaluate cost-effective treatment interventions.

However, it has been shown that preference-based instruments yield different utility values in the evaluation of the same patient population.[151] Marra et al. compared and determined the dimensions of health measured by four indirect utility instruments (the HUI2, the HUI3, the EQ-5D, and the SF-6D) and found that the scores derived using each of the four instruments were statistically and clinically different.[88] The results from the present study are consistent with these findings and it is therefore unlikely that the WOMAC derived HUI3 and the SF-6D would lead to comparable estimates as the weights for quality-adjusted life years (QALYs). As a result, an evaluation of the cost-effectiveness of UKA and TKA may vary depending on the utility scores used in the analysis. The HUI3 demonstrated greater improvement at 3 and 6 months following surgery in terms of the mean scores at both assessment times and would therefore yield a greater QALY compared to that of the SF-6D. For example, in a hypothetical situation of 100 patients who underwent UKA with a cost of $5000 per procedure and a mean change score from baseline to six months in the HUI3 and SF-6D of 0.24 and 0.13, respectively compared to

100 patients that underwent TKA with a cost of $4000 per surgery and a mean change score from baseline to six months in the HUI3 and SF-6D of 0.23 and 0.14, respectively; depending on which utility measure was chosen either treatment could be considered more cost-effective than the other. If the HUI3 was used as the weight for the QALY in the analysis, UKA would be deemed to have more value for money, whereas if the SF-6D were used TKA would be the more cost-effective treatment.

5.3 STUDY STRENGTHS AND LIMITATIONS

5.3.1 Strengths

This was a well designed study with large sample sizes thus enhancing the credibility of the findings. To evaluate the validity and responsiveness of the HUI3 and SF-6D preference based instruments we adapted methods used by Marra et al.[90] and we were therefore able to answer practically relevant questions such as: what magnitude of change in the SF-6D translates into a meaningful improvement in pain based on the percent change on the WOMAC pain score. In return, this offers a broader applicability and ease of understanding for surgeons and health care workers in terms of the discriminative performance of the HUI3 and SF-6D.

Moreover, unlike the majority of previously published studies evaluating the effectiveness of knee arthroplasty, both disease-specific and generic instruments were used as outcomes in the assessment of UKA and TKA for the treatment of OA. Frequently, studies have reported one dimensional aspects of knee pain and function or survival of the prosthesis. However, HRQL instruments are used with increasing frequency in the evaluation of the treatment for chronic diseases such as OA because of their ease of administration and ability to

capture a more complete picture of patient's change in health making this study a more representative evaluation of UKA from the patient's point of view. In addition, the hierarchical linear model used in this analysis was able to account for the correlation inherent in the longitudinal data while controlling for the potential variability among patients nested within each surgeon as well as the variability between each of the five surgeons, thus adding credibility to this study.

In contrast to the other previously published studies that evaluated the effectiveness of UKA versus TKA using patient self-assessed HRQL instruments, comparability both demographically and with regards to disease severity based on preoperative general and osteoarthritis-specific HRQL scores was confirmed preoperatively. This is an important feature given that although randomization was not performed during the patient recruitment phase, disease-severity is assumed to be similar between the two treatment groups.

5.3.2 Limitations

The analysis for this study was performed on secondary data and for this reason body mass index, other co-morbid conditions and the date of onset of OA, which are all factors in the progression of the disease, was not available. Patients were not randomized into treatment groups and as a result although those that underwent UKA were eligible for both treatments it cannot be assumed that all patients that underwent TKA suffered from unicompartmental OA and thus were also eligible for UKA. However, there were no significant differences between the two groups with regards to HRQL scores at baseline indicating that patients who underwent a UKA suffered from similar disease severity than those who underwent TKA.

Household income was not directly available and therefore had to be derived from the median estimates for each patient's dissemination area of living using Statistics Canada Postal Code Conversion File. As a result patients were categorized into one of four annual neighbourhood median income groups in order to determine if any differences exist with regards to HRQL. Preoperative HRQL was slightly worse in the lower income group however they were not clinically significant given that these differences were below the minimum 20% change in WOMAC pain scores and half the standard deviation of the mean change score for the OKS, which are considered clinically meaningful thresholds for both instruments.[125, 126] We also did not have access to the education level or the number of individuals within each household which can both have an impact on annual household income. However, use of income as a proxy for SES has been used in previous studies and we believe offers a valid representation of SES for each patient.[148, 149]

There are several different ways to impute data, varying imputation methods have been proposed based on the number of missing observations and the type of questionnaire being analyzed. For this study, patients for whom there were four or more missing responses for the SF-12 scores were excluded from the analysis. Where less than four responses were missing, the missing responses were imputed using the method proposed by Perneger et al.[112] When imputing the PCS and MCS scores the mean score of the missing question from the remaining sample was used in its place. For the WOMAC, if two or more pain items, both stiffness items, or four or more function items were missing, the patient was excluded from the analysis of that subscale. As recommended in the WOMAC user guide, where one pain, one stiffness, or one to three function items were omitted, the mean of the remaining responses for the specific subscale was used.[113] A similar method was used for the Oxford knee questionnaire in which the mean value

representing all the other responses was used when two or fewer questions were omitted; if more than two questions were omitted the patient was excluded from that analysis.[114] Less than 10 percent of the data was imputed and it did not have a significant effect on the results of the analyses.

In terms of determining the discriminative ability of the WOMAC derived HUI3 and SF-6D utilities, a dichotomous regression model was used, which assigns probabilities of patients being classified as improvement or not improvement based on an external predictor. The OKS and the WOMAC were both used for this purpose, given that both are disease-specific questionnaires and have both been validated for use in the evaluation of knee arthroplasty and the treatment of OA, respectively. However, work is currently underway to determine a minimal clinically important difference (MCID) for the OKS. A 20% change on the WOMAC pain score has been shown to represent a meaningful change in clinical symptoms of knee OA and was therefore used as one of the external predictors for evaluating the responsiveness of the SF-6D.[126] Alternatively, for the OKS, a sensitivity analysis was performed using a range of cut-points from 10% to 30%. It has been recommended that for the evaluation of knee arthroplasty, half the standard deviation of the mean change score be used and for this sample this translated into approximately a 10% change in OKS scores at the six month assessment period.[125] However, the 20% cut-point had the highest probability of correctly identifying both patient groups who had improved or not improved in terms of their percent change on the WOMAC pain scores, and was therefore used as the other external predictor of outcome for both the SF-6D and HUI3 utilities. Once a MCID is established for the OKS, we expect that it will not significantly deviate from this value given that the WOMAC has been thoroughly validated as a responsive and reliable measure of outcome following treatment of OA and is therefore a good representation of the probability of correctly categorizing patients as improved or not improved.

5.4 RECOMMENDATIONS

5.4.1 The use of UKA for the treatment of unicompartmental knee OA

Studies have shown that knee OA is often confined to the medial, lateral, or patelofemoral compartments, with 25% of patients suffering from OA that is isolated to the medial portion of the knee.[49, 68] However, due to earlier reports of significantly lower revision rates, TKA is predominantly chosen as the primary treatment for symptomatic OA and is still currently used in patients in which only one compartment of the knee is affected.[152] This study therefore lends support for the use of UKA in conjunction with its less invasive nature with greater bone stock preservation, faster recovery, a lower risk of complications and better restoration of knee kinematics.[78, 80, 153, 154] It is thus evident that given the extensive list of advantages to using UKA, the underutilization of this procedure partly stems from the lack of head to head comparisons properly evaluating the effectiveness of UKA versus TKA. This study, through comparable HRQL outcomes between the two procedures, highlights the knowledge pertaining to recent advancements in prosthesis design and the identification of specific indications for successful outcomes following UKA.

The Biomet Oxford® unicompartmental implant (Warsaw, Indiana) has demonstrated 10 year prosthetic survival rates as high as 98% in part by using the following criteria for inclusion: disabling knee pain; a non-inflammatory arthropathy in which the full thickness of cartilage was preserved in the lateral compartment and confirmed by preoperative radiographs; a functionally intact anterior cruciate ligament; fixed flexion deformity less than 15°.[53] The use of UKA as the primary treatment for unicompartmental arthroplasty is thus warranted and can now be used more frequently, without hesitation, for patients presenting with the appropriate indications.

5.4.2 Further research

Now that the WOMAC derived HUI3 has been evaluated in terms of its validity and responsiveness following knee arthroplasty, further study is needed to determine whether the WOMAC derived HUI3 provides the same result as the HUI3 itself. The lack of reproducibility between preference based instruments has been shown from previous studies and this study was no exception. However, further research is needed to test the reliability of the WOMAC derived HUI3 and the HUI3 itself in the same patient population to assess their similarity in terms of detecting change in scores following knee arthroplasty. Once this is determined, the WOMAC derived HUI3 can be used and implemented in the evaluation of knee arthroplasty for the treatment of OA as well as in other disease states without hesitation if proved to be valid.

Secondly, the primary purpose of accruing this data sample, which incorporated a new protocol in an attempt to increase the volume of knee arthroplasty procedures, maintain a high quality of care and decrease costs, has not yet been compared to patients undergoing surgery before the implementation of this model. The results of the protocol to increase efficiency and decrease hospital stay were successful and are consistent with a previous study put forth by Lombardi et al.[72] who sought to compare the incidence of complications, recovery, post-operative function and return to sport and work between minimally invasive UKA and less invasive TKA, which uses a smaller incision and does not require the removal of the patella during surgery, using an identical rapid recovery protocol for both groups. Lombardi et al. found that the faster a patient recovered, the lesser the burden placed on the overall healthcare system and when applied to patients with the appropriate indications, UKA demonstrated a more rapid recovery compared to TKA. However, the evaluation of the patient self-reported disease-specific

and generic HRQL instruments used in this study need to be compared to patients who underwent knee arthroplasty outside of the umbrella of the this efficiency pathway to determine if a rapid recovery protocol in fact translates into identical or improved outcomes following both UKA and TKA.

Moreover, patients undergoing knee arthroplasty are interested in knowing the probability of experiencing a meaningful change in health for the risk that is inherent in the intervention which can include: deep vein thrombosis, aseptic loosening, infection, and loss of mobility. This is specifically relevant with elective procedures such as UKA and TKA. The concept of a minimal clinically important difference (MCID) is therefore a key component in the assessment of patient-related outcomes. Frequently, statistical significance does not translate into clinical significance. Given that the assessment of pain is of primary interest and it can only be evaluated through patient self-reported instruments, a MCID which answers the question from the patient's perspective: "Is this change meaningful to me?" is highly relevant in clinical practice.[155]

Minimal clinically important differences are calculated using anchor-based methods where a global question is posed about the overall change in health between two assessment times which is then used as an anchor. This minimal change on the global scale (usually the "somewhat better" response) is used to calculate the corresponding improvement on a disease-specific scale.[155] If the present study had collected a general question on the overall health of each patient, then these data could have been used to calculate a MCID for the OKS. Therefore a study of this magnitude is needed with access to a global question of health in order to calculate this necessary threshold in patient-reported outcomes.

5.5 CONCLUSIONS

Unicompartmental knee arthroplasty as a treatment for OA affecting one compartment of the knee has and continues to evolve since its introduction almost forty years ago. However, scepticism remains due to the paucity of well controlled studies evaluating HRQL outcomes following surgery. In addition, previous studies have used a range of outcome measures thus making it difficult to compare results due to the variability in scores and lack of interpretability. This study therefore quells these shortcomings and fills in the gaps in the present literature creating a greater potential for the increased use of UKA. The WOMAC is the most commonly used psychometric HRQL instrument in the evaluation of patient-reported outcomes in knee arthroplasty and therefore demonstrating the validity and responsiveness of the WOMAC derived HUI3 along with the frequently used SF-6D contributes greatly to the use of generic utility measures in the evaluation of knee arthroplasty. This is important given their broad use of applicability and will allow for the comparison across different interventions for use as weights in QALYs for cost-effectiveness analyses.

With the implementation of validated and responsive generic utility measures as tools for the evaluation of HRQL outcomes following surgery relevant patient-specific factors can be identified. Socioeconomic status is one factor that is well known to effect treatment outcomes for many disease states. However, no studies have been performed which look to see if differences in outcome between UKA and TKA exist between SES groups. Using income as a proxy for SES, it does not appear to have a significant impact on pre-operative scores or on HRQL following knee arthroplasty in this patient sample. Moreover, age and gender have also been previously identified as having significant effects on patient outcomes following surgery and

although in this study age was not a significant covariate in the model; males had significantly better total WOMAC and OKS scores at six months.

REFERENCES

1. Canadian Arthritis Network. Partnerships: Strategic research initiative in osteoarthritis. Available at: www.arthritisnetwork.ca.

2. Arden N, Cooper C. The background of osteoarthritis. In: Arden N, Cooper C, eds. *Osteoarthritis Handbook.* First ed. New York, NY: Taylor & Francis; 2006:3-24.

3. Buckwalter JA. Osteoarthritis and articular cartilage use, disuse, and abuse: Experimental studies. *Journal of Rheumatology - Supplement*. 1995;43:13-15.

4. President and Fellows of Harvard College. Osteoarthritis: Causes and diagnosis. *Harvard Men's Health Watch.* 2006;10:1.

5. Iorio R, Robb WJ, Healy WL, et al. Orthopaedic surgeon workforce and volume assessment for total hip and knee replacement in the united states: Preparing for an epidemic. *J Bone Joint Surg Am*. 2008;90:1598-1605.

6. Escobar A. Development of explicit criteria for total knee replacement. *Int J Technol Assess Health Care*. 2003;19:57.

7. Arden, E., Arden, N., Hunter, D. *Osteoarthritis: The Facts.* First ed. New York: Oxford University Press; 2008.

8. Anderson JJ, Felson DT. Factors associated with osteoarthritis of the knee in the first national health and nutrition examination survey (HANES I). evidence for an association with overweight, race, and physical demands of work. *Am J Epidemiol*. 1988;128:179-189.

9. Altman R, Asch E, Bloch D, Bole G, Broenstein D, Brandt K. The american college of rheumatology criteria for the classification and reporting of osteoarthritis of the knee. *Arthritis Rheum*. 1986:1039-1049.

10. Brandt K. *An Atlas of Osteoarthritis.* First ed. New York, NY.: Parthenon Pub. Group; 2001.

11. Felson DT. Osteoarthritis. *Rheum Dis Clin North Am*. 1990;16:499-512.

12. Scott RD. *Total Knee Arthroplasty.* First ed. Philadelphia, Pennsylvania: Saunders Elsevier; 2006.

13. Canadian Institute for Health Information. Canadian joint replacement registry. *CJRR*. 2007;Hip and Knee Replacements in Canada:1.

14. Riddle DL, Jiranek WA, McGlynn FJ. Yearly incidence of unicompartmental knee arthroplasty in the united states. *J Arthroplasty*. 2008;23:408-412.

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;89:780-785.

16. DeFrances CJ, Lucas CA, Buie VC, Golosinskiy A. 2006 national hospital discharge survey. *National health statistics reports*. 2008:1-20.

17. Fitz W. Unicompartmental knee arthroplasty with use of novel patient-specific resurfacing implants and personalized jigs. *J Bone Joint Surg Am*. 2009;91:69-76.

18. Archibeck MJ, White RE,Jr. What's new in adult reconstructive knee surgery. *J Bone Joint Surg Am*. 2006;88:1677-1686.

19. Bettinson KA, Pinder IM, Moran CG, Weir DJ, Lingard EA. All-polyethylene compared with metal-backed tibial components in total knee arthroplasty at ten years. A prospective, randomized controlled trial. *J Bone Joint Surg Am*. 2009;91:1587-1594.

20. Tsao AK. Jones LC. Lewallen DG. What patient and surgical factors contribute to implant wear and osteolysis in total joint arthroplasty? *J Am Acad Orthop Surg*. 2008;16:S7-13.

21. Gioe TJ, Stroemer ES, Santos ER. All-polyethylene and metal-backed tibias have similar outcomes at 10 years: A randomized level I [corrected] evidence study. *Clin Orthop Relat Res*. 2007;455:212-218.

22. 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;18:9-15.

23. Rand JA. Factors affecting the durability of primary total knee prostheses. *J Bone Joint Surg Am*. 2003;85-A:259-65.

24. Rodriguez JA, Baez N, Rasquinha V, Ranawat CS. Metal-backed and all-polyethylene tibial components in total knee replacement. *Clin Orthop Relat Res*. 2001:174-183.

25. Udomkiat P, Dorr LD, Long W. Matched-pair analysis of all-polyethylene versus metal-backed tibial components. *J Arthroplasty*. 2001;16:689-696.

26. Gioe TJ, Glynn J, Sembrano J, Suthers K, Santos ERG, Singh J. Mobile and fixed-bearing (all-polyethylene tibial component) total knee arthroplasty designs. A prospective randomized trial. *J Bone Joint Surg Am*. 2009;91:2104-2112.

27. Cuckler JM, Lemons J, Tamarapalli JR, Beck P. Polyethylene damage on the nonarticular surface of modular total knee prostheses. *Clin Orthop Relat Res*. 2003:248-253.

28. Engh G, Dwyer K, Hanes C. Polyethylene wear of metal-backed tibial components in total and unicompartmental knee prostheses. *J Bone Joint Surg Br*. 1992;74-B:9-17.

29. Jones SM, Pinder IM, Moran CG, Malcolm AJ. Polyethylene wear in uncemented knee replacements. *J Bone Joint Surg Br*. 1992;74:18-22.

30. Colizza W, Insall J, Scuderi G. The posterior stabilized total knee prosthesis. assessment of polyethylene damage and osteolysis after a ten-year-minimum follow-up. *J Bone Joint Surg Am*. 1995;77:1713-1720.

31. Bryan RS, Rand JA. Revision total knee arthroplasty. *Clin Orthop Relat Res*. 1982:116-122.

32. Dieppe P. Knee replacement surgery for osteoarthritis: Effectiveness, practice variations, indications and possible determinants of utilization. *Rheumatol*. 1999;38:73-83.

33. Back DL, Cannon SR, Hilton A, Bankes MJ, Briggs TW. The kinemax total knee arthroplasty. nine years' experience. *J Bone Joint Surg Br*. 2001;83:359-363.

34. Ewald FC, Wright RJ, Poss R, Thomas WH, Mason MD, Sledge CB. Kinematic total knee arthroplasty: A 10- to 14-year prospective follow-up review. *J Arthroplasty*. 1999;14:473-480.

35. Martin SD, McManus JL, Scott RD, Thornhill TS. Press-fit condylar total knee arthroplasty. 5- to 9-year follow-up evaluation. *J Arthroplasty*. 1997;12:603-614.

36. Rorabeck CH, Murray P. The cost benefit of total knee arthroplasty. *Orthopedics*. 1996;19:777-779.

37. Falatyn S, Lachiewicz PF, Wilson FC. Survivorship analysis of cemented total condylar knee arthroplasty. *Clin Orthop Relat Res*. 1995:178-184.

38. Malkani A, Rand J, Bryan R, Wallrichs S. Total knee arthroplasty with the kinematic condylar prosthesis. A ten-year follow-up study. *J Bone Joint Surg Am*. 1995;77:423-431.

39. Kirwan JR, Currey HL, Freeman MA, Snow S, Young PJ. Overall long-term impact of total hip and knee joint replacement surgery on patients with osteoarthritis and rheumatoid arthritis. *Br J Rheumatol*. 1994;33:357-360.

40. Ranawat CS. Long-term results of the total condylar knee arthroplasty. A 15-year survivorship study. *Clin Orthop Relat Res*. 1993:94-102.

41. Gill GS, Mills DM. Long-term follow-up evaluation of 1000 consecutive cemented total knee arthroplasties. *Clin Orthop Relat Res*. 1991:66-76.

42. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster J. Health-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;86:963-974.

43. Callahan CM, Drake BG, Heck DA, Dittus RS. Patient outcomes following tricompartmental total knee replacement. A meta-analysis. *J Am Med Assoc*. 1994;271:1349-1357.

44. Norman-Taylor FH, Palmer CR, Villar RN. Quality-of-life improvement compared after hip and knee replacement. *J Bone Joint Surg Br*. 1996;78:74-77.

45. Rissanen P, Aro S, Sintonen H, Slatis P, Paavolainen P. Quality of life and functional ability in hip and knee replacements: A prospective study. *Qual Life Res*. 1996;5:56-64.

46. Wright RJ, Sledge CB, Poss R, Ewald FC, Walsh ME, Lingard EA. Patient-reported outcome and survivorship after kinemax total knee arthroplasty. *J Bone Joint Surg Am*. 2004;86:2464-2470.

47. Deshmukh RV. Unicompartmental knee arthroplasty: Long-term results. *Clin Orthop Relat Res*. 2001:272-8.

48. Weale AE, Murray DW, Crawford R, et al. Does arthritis progress in the retained compartments after 'oxford' medial unicompartmental arthroplasty?: A clinical and radiological study with a minimum ten-year follow-up. *J Bone Joint Surg Br*. 1999;81-B:783-789.

49. Ackroyd CE. Medial compartment arthroplasty of the knee. *J Bone Joint Surg Br*. 2003;85-B:937-942.

50. 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:1126-1130.

51. Murray DW, Goodfellow JW, O'Connor JJ. The oxford medial unicompartmental arthroplasty: A ten-year survival study. *J Bone Joint Surg Br*. 1998;80-B:983-989.

52. Lewold S. Oxford meniscal bearing knee versus the marmor knee in unicompartmental arthroplasty for arthrosis. A swedish multicenter survival study. *J Arthroplasty*. 1995;10:722-31.

53. Price AJ. Long-term clinical results of the medial oxford unicompartmental knee arthroplasty. *Clin Orthop Relat Res*. 2005:171-80.

54. Cartier P. Unicompartmental knee arthroplasty surgery. 10-year minimum follow-up period. *J Arthroplasty*. 1996;11:782-8.

55. Emerson RH. Comparison of a mobile with a fixed-bearing unicompartmental knee implant. *Clin Orthop Relat Res*. 2002:62-70.

56. Price AJ, Short A, Kellett C, et al. Ten-year in vivo wear measurement of a fully congruent mobile bearing unicompartmental knee arthroplasty. *J Bone Joint Surg Br*. 2005;87-B:1493-1497.

57. Emerson RH, Jr., Higgins LL. Unicompartmental knee arthroplasty with the oxford prosthesis in patients with medial compartment arthritis. *J Bone Joint Surg Am*. 2008;90:118-122.

58. Engelbrecht E, Siegel A, Rottger J, Buchholz HW. Statistics of total knee replacement: Partial and total knee replacement, design st. georg: A review of a 4-year observation. *Clin Orthop Relat Res*. 1976:54-64.

59. Insall J, Aglietti P. A five to seven-year follow-up of unicondylar arthroplasty. *J Bone Joint Surg Am*. 1980;62:1329-1337.

60. Wilcox PG, Jackson DW. Unicompartmental knee arthroplasty. *Orthop Rev*. 1986;15:490-495.

61. Kozinn SC, Scott R. Unicondylar knee arthroplasty. *J Bone Joint Surg Am*. 1989;71:145-150.

62. Padgett DE, Stern SH, Insall JN. Revision total knee arthroplasty for failed unicompartmental replacement. *J Bone Joint Surg Am*. 1991;73:186-190.

63. Mallory TH, Danyi J. Unicompartmental total knee arthroplasty. A five- to nine-year follow-up study of 42 procedures. *Clin Orthop Relat Res*. 1983:135-138.

64. Berger RA. Unicompartmental knee arthroplasty: Clinical experience at 6- to 10-year followup. *Clin Orthop Relat Res*. 1999;367:50-60.

65. Rajasekhar 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:983-985.

66. Romanowski MR, Repicci JA. Minimally invasive unicondylar arthroplasty: Eight-year follow-up. *J Knee Surg*. 2002;15:17-22.

67. Price AJ, Webb J, Topf H, et al. Rapid recovery after oxford unicompartmental arthroplasty through a short incision. *J Arthroplasty*. 2001;16:970-976.

68. Iorio R, Healy WL. Unicompartmental arthritis of the knee. *J Bone Joint Surg Am*. 2003;85:1351-1364.

69. Argenson JA, Chevrol-Benkeddache Y, Aubaniac J. Modern unicompartmental knee arthroplasty with cement: A three to ten-year follow-up study. *J Bone Joint Surg Am*. 2002;84:2235-2239.

70. 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 50,493 knee replacements from the finnish arthroplasty register. *Acta orthop*. 2008;79:499-507.

71. Amin AK. Unicompartmental or total knee arthroplasty?: Results from a matched study. *Clin Orthop Relat Res*. 2006;451:101-6.

72. Lombardi AV,Jr, Berend KR, Walter CA, Aziz-Jacobo J, Cheney NA. Is recovery faster for mobile-bearing unicompartmental than total knee arthroplasty?. *Clin Orthop Relat Res*. 2009;467:1450-1457.

73. Hopper GP, Leach WJ. Participation in sporting activities following knee replacement: Total versus unicompartmental. *Knee Surg Sports Traumtol Arthrosc*. 2008;16:973-979.

74. Manzotti A, Confalonieri N, Pullen C. Unicompartmental versus computer-assisted total knee replacement for medial compartment knee arthritis: A matched paired study. *Int Orthop*. 2007;31:315-319.

75. Hassaballa MA, Porteous AJ, Learmonth ID. Functional outcomes after different types of knee arthroplasty: Kneeling ability versus descending stairs. *Med Sci Monit*. 2007;13:R77-81.

76. Yang KY, Wang MC, Yeo SJ, Lo NN. Minimally invasive unicondylar versus total condylar knee arthroplasty--early results of a matched-pair comparison. *Singapore Med J*. 2003;44:559-562.

77. Weale AE, Halabi OA, Jones PW, White SH. Perceptions of outcomes after unicompartmental and total knee replacements. *Clin Orthop Relat Res*. 2001:143-153.

78. Robertsson O, Borgquist L, Knutson K, Lewold S, Lidgren L. Use of unicompartmental instead of tricompartmental prostheses for unicompartmental arthrosis in the knee is a cost-effective alternative. 15,437 primary tricompartmental prostheses were compared with 10,624 primary medial or lateral unicompartmental prostheses. *Acta Orthop Scand*. 1999;70:170-175.

79. Laurencin CT. Unicompartmental versus total knee arthroplasty in the same patient. A comparative study. *Clin Orthop Relat Res*. 1991:151-6.

80. Rougraff BT. A comparison of tricompartmental and unicompartmental arthroplasty for the treatment of gonarthrosis. *Clin Orthop Relat Res*. 1991:157-64.

81. Newman JH. Unicompartmental or total knee replacement? five-year results of a prospective, randomised trial of 102 osteoarthritic knees with unicompartmental arthritis. *J Bone Joint Surg Br*. 1998;80:862-5.

82. Newman JH, 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:52-57.

83. Walton NP, Jahromi I, Lewis PL, Dobson PJ, Angel KR, Campbell DG. Patient-perceived outcomes and return to sport and work: TKA versus mini-incision unicompartmental knee arthroplasty. *J Knee Surg*. 2006;19:112-116.

84. Wolfe F. Determinants of WOMAC function, pain and stiffness scores: Evidence for the role of low back pain, symptom counts, fatigue and depression in osteoarthritis, rheumatoid arthritis and fibromyalgia. *Rheumatol*. 1999;38:355-61.

85. Bellamy N. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. consensus development at OMERACT III. *J Rheumatol*. 1997;24:799-802.

86. 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-B:63-69.

87. Khanna D. Health-related quality of life--an introduction. *Am J Manag Care.* 2007;13 Suppl 9:S218.

88. Marra CA. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care*. 2004;42:1125-31.

89. Devlin N. Does NICE have a cost-effective threshold and what other factors influence its decisions? A binary choice analysis. *J Health Econ*. 2004;13:437-52.

90. Marra CA, Rashidi AA, Guh D, et al. Are indirect utility measures reliable and responsive in rheumatoid arthritis patients?. *Qual Life Res*. 2005;14:1333-1344.

91. McTaggart-Cowan HM, Marra CA, Yang Y, et al. The validity of generic and condition-specific preference-based instruments: The ability to discriminate asthma control status. *Qual Life Res*. 2008;17:453-462.

92. Fortin PR. Timing of total joint replacement affects clinical outcomes among patients with osteoarthritis of the hip or knee. *Arthitis Rheum*. 2002;46:3327-30.

93. Lingard EA, Katz JN, Wright EA, Sledge CB, the Kinemax Outcomes Group . Predicting the outcome of total knee arthroplasty. *J Bone Joint Surg Am*. 2004;86:2179-2186.

94. Sharma L, Sinacore J, Daugherty C, et al. Prognostic factors for functional outcome of total knee replacement: A prospective study. *J Gerontol Med Sci*. 1996;51:M152-7.

95. Mahomed NN, Liang MH, Cook EF, et al. The importance of patient expectations in predicting functional outcomes after total joint arthroplasty. *J Rheumatol*. 2002;29:1273-1279.

96. Yong PF, Milner PC, Payne JN, Lewis PA, Jennison C. Inequalities in access to knee joint replacements for people in need. *Ann Rheum Dis*. 2004;63:1483-1489.

97. Yoo HY, Thuluvath PJ. Outcome of liver transplantation in adult recipients: Influence of neighborhood income, education, and insurance. *Liver Transpl*. 2004;10:235-243.

98. Rovelli M, Palmeri D, Vossler E, Bartus S, Hull D, Schweizer R. Noncompliance in renal transplant recipients: Evaluation by socioeconomic groups. *Transplant Proc*. 1989;21:3979-3981.

99. Davis ET, Lingard EA, Schemitsch EH, Waddell JP. Effects of socioeconomic status on patients' outcome after total knee arthroplasty. *Int J Qual Health Care*. 2008;20:40-46.

100. Côté I. Quality of life in hypertension: The SF-12 compared to the SF-36. *Can J Clin Pharmacol*. 2004;11:e232-8.

101. Wee CC. Comparing the SF-12 and SF-36 health status questionnaires in patients with and without obesity. *Health Qual Life Outcomes*. 2008;6:11-18.

102. Kopec JA, Richardson CG, Llewellyn-Thomas H, Klinkhoff A, Carswell A, Chalmers A. Probabilistic threshold technique showed that patients' preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied. *J Clin Epidemiol*. 2007;60:929-938.

103. Grootendorst P. Health utilities index mark 3: Evidence of construct validity for stroke and arthritis in a population health survey. *Med Care*. 2000;38:290-9.

104. Grootendorst P, Marshal D, Pericak D, Bellamy N, Feeny D, Torrance GW. A model to estimate health utiities index mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the knee. *J Rheumatol*. 2007;34:534-542.

105. Walters SJ. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res*. 2005;14:1523-32.

106. de Vries M. Comparison of generic and disease-specific questionnaires for the assessment of quality of life in patients with peripheral arterial disease. *J Vasc Surg*. 2005;41:261-8.

107. Bellamy N. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15:1833-40.

108. Bellamy N. WOMAC: A 20-year experiential review of a patient-centered self-reported health status questionnaire. *J Rheumatol*. 2002;29:2473-6.

109. Söderman P. Validity and reliability of swedish WOMAC osteoarthritis index: A self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). *Acta orthop scand*. 2000;71:39-46.

110. Garratt AM, Brealey S, Gillespie WJ. Patient-assessed health instruments for the knee: A structured review. *Rheumatol*. 2004;43:1414-1423.

111. Sanderson K. The SF-12 in the australian population: Cross-validation of item selection. *Aust N Z J Public Health*. 2002;26:343-5.

112. Perneger TV, Burnand B. A simple imputation algorithm reduced missing data in SF-12 health surveys. *J Clin Epidemiol*. 2005;58:142-149.

113. Bellamy N. *WOMAC Osteoarthritis Index User Guide. Version IV.* Queensland: Australia: ; 2003.

114. Murray DW, Fitzpatrick R, Rogers K, et al. The use of the oxford hip and knee scores. *J Bone Joint Surg Br*. 2007;89-B:1010-1014.

115. Marshall D. Validation of a prediction model to estimate health utilities index mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the hip. *Value Health*. 2008;11:470-7.

116. Brazier JE. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42:851-9.

117. Shrout PE FJ. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;2:420-428.

118. Cohen J, ed. *Statistical Power Analysis for the Behavioral Sciences.* Second ed. New York: Academic Press; 1978.

119. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin Ther*. 1996;18:979-992.

120. Liang MH. Evaluating measurement responsiveness. *J Rheumatol*. 1995;22:1191-1192.

121. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2:1-74.

122. Cohen J. A power primer. *Psychol Bull*. 1992:155-159.

123. Blanchard C, Feeny D, Mahon JL, et al. Is the health utilities index valid in total hip arthroplasty patients? *Qual Life Res*. 2004;13:339-348.

124. Crawley MJ. The R book. In: Wiley J, ed. *The R Book.* First ed. England: Springer; 2007:47-322.

125. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Med Care*. 2003;41:582-592.

126. Bellamy N, Bell MJ, Goldsmith CH, et al. Evaluation of WOMAC 20, 50, 70 response criteria in patients treated with hylan G-F 20 for knee osteoarthritis. *Ann Rheum Dis*. 2005;64:881-885.

127. Liang MH. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. 1990;28:632-42.

128. Marx RG. Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome. *J Bone Joint Surg Am*. 2005;87:1999-2005.

129. Bombardier C. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care*. 1995;33:AS131-44.

130. Beaton DE, Richards RR. Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. *J Bone Joint Surg Am*. 1996;78:882-90.

131. Pandit H. The oxford medial unicompartmental knee replacement using a minimally-invasive approach. *J Bone Joint Surg Br*. 2006;88:54-60.

132. Li G, Suggs J, Hanson G, Durbhakula S, Johnson T, Freiberg A. Three-dimensional tibiofemoral articular contact kinematics of a cruciate-retaining total knee arthroplasty. *J Bone Joint Surg Am*. 2006;88:395-402.

133. Stern S, Insall J. Posterior stabilized prosthesis. results after follow-up of nine to twelve years. *J Bone Joint Surg Am*. 1992;74:980-986.

134. Beard DJ. The influence of the presence and severity of pre-existing patellofemoral degenerative changes on the outcome of the oxford medial unicompartmental knee replacement. *J Bone Joint Surg Br*. 2007;89:1597-1601.

135. Gartland JJ. Orthopaedic clinical research. deficiencies in experimental design and determinations of outcome. *J Bone Joint Surg Am*. 1988;70:1357-1364.

136. Gill GS. Total condylar knee arthroplasty. 16- to 21-year results. *Clin Orthop Relat Res* [serial online]. 1999:210.

137. Ackroyd CE. A comparative study of the medial st georg sled and kinematic total knee arthroplasties. ten-year survivorship. *J Bone Joint Surg Br* [serial online]. 2002;84:667.

138. Kurtz S. Prevalence of primary and revision total hip and knee arthroplasty in the united states from 1990 through 2002. *Journal of bone and joint surgery. American volume* [serial online]. 2005;87:1487.

139. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J Educ Behav Stats*. 1998;24:323-355.

140. Hawker G, Wright J, Coyte P, et al. Health-related quality of life after knee replacement. results of the knee replacement patient outcomes research team study. *J Bone Joint Surg Am*. 1998;80:163-73.

141. Gittelsohn AM, Halpern J, Sanchez RL. Income, race, and surgery in maryland. *Am J Public Health*. 1991;81:1435-1441.

142. Humphries KH, van Doorslaer E. Income-related health inequality in canada. *Soc Sci Med*. 2000;50:663-671.

143. Hegewald MJ, Crapo RO. Socioeconomic status and lung function. *Chest*. 2007;132:1608-1614.

144. Clark AM, DesMeules M, Luo W, Duncan AS, Wielgosz A. Socioeconomic status and cardiovascular disease: Risks and implications for care. *Nature Reviews Cardiology*. 2009;6:712-722.

145. Grotto I, Huerta M, Sharabi Y. Hypertension and socioeconomic status. *Curr Opin Cardiol*. 2008;23:335-339.

146. Robertsson O. The routine of surgical management reduces failure after unicompartmental knee arthroplasty. *Journal of bone and joint surgery* [serial online]. 2001;83:45.

147. SooHoo NF, Lieberman JR, Ko CY, Zingmond DS. Factors predicting complication rates following total knee replacement. *J Bone Joint Surg Am*. 2006;88:480-485.

148. Marra CA, Lynd LD, Esdaile JM, Kopec J, Anis AH. The impact of low family income on self-reported health outcomes in patients with rheumatoid arthritis within a publicly funded health-care environment. *Rheumatology*. 2004;43:1390-1397.

149. Lynd LD, Sandford AJ, Kelly EM, et al. Reconcilable differences: A cross-sectional study of the relationship between socioeconomic status and the magnitude of short-acting beta-agonist use in asthma. *Chest*. 2004;126:1161-1168.

150. Kozyrskyj AL, Kendall GE, Jacoby P, Sly PD, Zubrick SR. Association between socioeconomic status and the development of asthma: Analyses of income trajectories. *Am J Public Health*. 2010;100:540-546.

151. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med Care*. 2003;41:791-801.

152. Knutson K, Robertsson O,Lidgren. The swedish knee arhtroplasty register. A nation-wide study of 30 003 knees 1976-1992. *Acta Orthop Scand*. 1994:375-86.

153. Waite JC, Dodd CA, Murray DW, Gill HS. In vivo kinematics of the ACL-deficient limbduring running and cutting. *Knee Surg Sports Traumtol Arthrosc*. 2005:377-84.

154. Goodfellow JW, Kershaw CJ, D'A B,M.K., O'Connor JJ. The oxford knee for unicompartmental osteoarthritis. *J Bone Joint Surg Br*. 1988:692-701.

155. Singh J, Sloan JA, Johanson NA. Challenges with health-related quality of life assessment in arthroplasty patients: Problems and solutions. *J Am Acad Orthop Surg*. 2010;18:72-82.

APPENDIX I – ETHICS APPROVAL FORMS

Ethics approval was received from the University of British Columbia Research Ethics Board (#C04-0577) and research approval was received from the Richmond Health Services Research Advisory Committee.

The University of British Columbia
Office of Research Services,
Clinical Research Ethics Board – Room 210, 828 West 10th Avenue, Vancouver, BC V5Z 1L8

Certificate of Expedited Approval

Clinical Research Ethics Board Official Notification

PRINCIPAL INVESTIGATOR	DEPARTMENT	NUMBER
Roberts, Cindy		C04-0577

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT

Vancouver Coastal Health Authority

CO-INVESTIGATORS:

Chambers, Keith, Health Care/Epidemiology; Chao, Carole, ; Coleman, Jeff, ; De Paoli, Lisa, ; Dong, Glenn, ; Fenton, Jonathon, ; Fitzgerald, Mark, Medicine; Gillam, Carole, ; Gilliver, Andrea, ; Hughes, Kenneth; Keyes, Nancy; Latteier, Jim; Magri, Laurel; Masson, Becky; Palmer, Tarina; Pardy, Petra; Sainas, Stephanie; Schmidt, Brian; Sobolev, Boris; Tsou, Patsy; Varona, Fabiola

SPONSORING AGENCIES

Provincial Health Services Authority

TITLE :

Arthroplasty Enhancement Pilot

APPROVAL DATE	TERM (YEARS)	DOCUMENTS INCLUDED IN THIS APPROVAL:
3 November 2004	1	Protocol Version 5a, dated 27 September 2004; Letter of Invitation Version 1, dated 22 October 2004; Patient Information and Consent Form, Version 3, dated 1 November, 2004; WOSFOX HIP and KNEE Questionnaires; Priority Criteria Tool Revised April 2003;

CERTIFICATION:

In respect of clinical trials:

1. The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards defined in Division 5 of the Food and Drug Regulations.

2. The Research Ethics Board carries out its functions in a manner consistent with Good Clinical Practices.

3. This Research Ethics Board has reviewed and approved the clinical trial protocol and informed consent form for the trial which is to be conducted by the qualified investigator named above at the specified clinical trial site. This approval and the views of this Research Ethics Board have been documented in writing.

The documentation included for the above-named project has been reviewed by the Chair of the UBC CREB, and the research study, as presented in the documentation, was found to be acceptable on ethical grounds for research involving human subjects and was approved by the UBC CREB.

The CREB approval for this study expires one year from the approval date.

Approval of the Clinical Research Ethics Board by one of:

Dr. P. Loewen, Chair

Dr. A. Gagnon, Associate Chair

Dr. J. McCormack, Associate Chair

The University of British Columbia
Office of Research Services,
Clinical Research Ethics Board – Room 210, 828 West 10th Avenue, Vancouver, BC V5Z 1L8

Certificate of Expedited Approval

Clinical Research Ethics Board Official Notification

PRINCIPAL INVESTIGATOR	DEPARTMENT	NUMBER
Roberts, Cindy		C04-0577

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT
Vancouver Coastal Health Authority

CO-INVESTIGATORS:
Chambers, Keith, Health Care/Epidemiology; Chao, Carole, ; Coleman, Jeff, ; De Paoli, Lisa, ; Dong, Glenn, ; Fenton, Jonathon, ; Fitzgerald, Mark, Medicine; Gillam, Carole, ; Gilliver, Andrea, ; Hughes, Kenneth; Keyes, Nancy; Latteier, Jim; Magri, Laurel; Masson, Becky; Palmer, Tarina; Pardy, Petra; Sainas, Stephanie; Schmidt, Brian; Sobolev, Boris; Tsou, Patsy; Varona, Fabiola

SPONSORING AGENCIES
Vancouver Coastal Health Authority (Ethics only)

TITLE :
Arthroplasty Enhancement Pilot

APPROVAL DATE	TERM (YEARS)	DOCUMENTS INCLUDED IN THIS APPROVAL:
3 November 2004	1	Protocol Version 5a, dated 27 September 2004; Letter of Invitation Version 1, dated 22 October 2004; Patient Information and Consent Form, Version 3, dated 1 November, 2004; WOSFOX HIP and KNEE Questionnaires; Priority Criteria Tool Revised April 2003;

CERTIFICATION:
In respect of clinical trials:
1. The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards defined in Division 5 of the Food and Drug Regulations.
2. The Research Ethics Board carries out its functions in a manner consistent with Good Clinical Practices.
3. This Research Ethics Board has reviewed and approved the clinical trial protocol and informed consent form for the trial which is to be conducted by the qualified investigator named above at the specified clinical trial site. This approval and the views of this Research Ethics Board have been documented in writing.

The documentation included for the above-named project has been reviewed by the Chair of the UBC CREB, and the research study, as presented in the documentation, was found to be acceptable on ethical grounds for research involving human subjects and was approved by the UBC CREB.

The CREB approval for this study expires one year from the approval date.

Approval of the Clinical Research Ethics Board by one of:
Dr. P. Loewen, Chair
Dr. A. Gagnon, Associate Chair
Dr. J. McCormack, Associate Chair