A NEW LOOK AT THE CROSS-SECTIONAL RELATIONSHIP OF SELF-REPORTED PAIN, FUNCTION AND WALKING PERFORMANCE WITH RADIOGRAPHIC WEAR AND OTHER EARLY INDICATORS OF TOTAL HIP REPLACEMENT FAILURE IN PATIENTS WITH OSTEOARTHRITIS

Jennifer Michelle Charlesworth

A thesis submitted in conformity with the requirements for the degree of Master of Science
Institute of Medical Science
University of Toronto

Copyright © Jennifer Michelle Charlesworth 2010
Abstract

Background: Total hip replacement (THR) with post-operative surveillance is recommended for debilitating osteoarthritis (OA). Using self-reported pain, function or walking performance is one alternative to address increasing surveillance demands.

Objective: A cross-sectional cohort study to evaluate the associations of pain, function and performance with two radiographic markers of potential THR failure.

Participants: 110 patients, median 6 years after THR surgery for OA.

Methods: Questionnaires assessed demographics, co-morbidity, arthritis severity, pain, pain catastrophizing, and functional status. Performance was measured using the six minute walk test. THR outcome was assessed radiographically.

Results: Few patients had pain, functional impairment or radiographic markers of potential THR failure. A larger percentage of patients with some intermittent pain (10.7 versus 8.6%) and pain
after walking performance (40.0 versus 27.6%) had higher wear, but these differences were not significant.

**Conclusion:** Measures of pain are potentially important for larger studies aiming to develop alternative methods of post-operative surveillance.
Acknowledgments

This work would not have been possible if not for the assistance of many individuals:

First and foremost I would like to thank my supervisor, Dr. Gillian Hawker, as she provided me with the opportunity to complete this degree. She has taken time from her busy schedule to guide me patiently through this process, provided copious feedback on various drafts of this manuscript, and has given me much insight into what makes a successful researcher, supervisor and physician. I strive to follow her example and I am grateful for her continued support.

Dr. Hans Kreder has provided valued leadership and mentorship. He has graciously given up hours of his time to assist with any dilemma, be it small or large. Moreover, he has taken the time to foster my learning beyond the boundaries of my research. For opportunities and support beyond price, I thank you.

I want now to acknowledge the remaining two members of my program advisory committee: Dr. Scott Thomas and Dr. Monique Gignac, who graciously stepped in to assist whenever needed. Scott volunteered to assist with supervision and has provided contacts, comments, and a sympathetic ear whenever required. Monique has reviewed many drafts, provided insightful feedback and met with me on multiple occasions to facilitate my work. I have learned from watching and listening to them work.

Next I want to acknowledge Dr. Mary Seeman who provided support throughout my studies in the IMS. She has acted as a mentor and fostered my efforts and varied interests. Moreover, she has taken an interest in my dreams for the future and facilitated them. I could not have asked for a better graduate coordinator for my program.

Similarly, all the staff at the Institute of Medical Science, present and past, must also be acknowledged for all their assistance.

I must thank Dr. Iris Weller for recognizing my interest and passion for clinical research. She took a chance on an unknown student and introduced me to most of the individuals noted above and below. Her years of work on the SafeT Study provided the basis for this project.

Monica Kunz and Katrine Milner have been valued colleagues throughout my degree. They have facilitated and assisted with multiple endeavours and made Sunnybrook a great place to complete my studies.

Dr. Cari Whyne provided access and equipment to facilitate my work. Without her assistance, obtaining x-rays necessary for this analysis would have been cost-prohibitive.

Dr. Fiona Webster was also always happy to assist. She provided tips for the writing process and met with me to help distil my ideas.
All the co-investigators on the SafeT Study have assisted in many ways. Dr. Jeffrey Gollish provided resources and suggestions to support this work. Dr. Steve MacDonald co-ordinated technical assistance and volunteered additional recruitment resources. I must especially acknowledge Dr. Paul Corey who took the time to explain some of the finer points of statistical analysis. I also must thank Deborah Kennedy for providing support, guidance, resources and a couple of, much needed, discussions. It has been a genuine pleasure working with all of these excellent researchers and clinicians.

My lab mates and colleagues stepped in to help on numerous occasions, even if it was just to listen to me think out loud. Jason, Diane, Carla, Madeline, Rebecca, Jennifer L, Kristin - I appreciate all that you did.

I must take a minute to acknowledge the hard work of Matthew Kennedy and Heeba Abdullah. These extra-ordinary individuals took time out their own busy lives to spend hours to assist with downloading x-rays and cleaning data.

Many thanks to all the surgeons and advanced practice physiotherapists, who tolerated a researcher invading their clinics. All the front line staff working in the Sunnybrook Musculoskeletal Program made the data collection process both possible and enjoyable. Special mention to the ladies in X-Ray: Lynn and Monique, the ladies in OPD: especially Cheryl, Alice and Judy, and the staff of HDR: David, Diane, and Lane.

To my colleagues at Women’s College Hospital, your willing assistance was always appreciated: Angela, Yalnee, Samara, Melissa - it has been a pleasure working with you. Jacob, in particular, has been both a valued colleague and a friend. He provided many useful suggestions, consulted with me on numerous occasions on anything related to statistics, and was always around to sympathize with coding issues.

Kory Charron at London Health Sciences center provided the wear analysis data. Dr. Shadi Shihata evaluated over 300 films for markers of hip replacement failure. Without these individuals I would not have had data to analyze.

Numerous friends have provided me with invaluable support. They provided a willing and sympathetic ear when things were not going well, a joyous cheer when things were, and some much needed fun at any time. You are too numerous to name but you all know who you are. Special mention goes out to Clarissa, my cheerleader and constant supporter; Amanda R, Amanda C, Mike L, Kathy M, Jennifer L, Myriam W, Pam P, Julie B and David B.

I must also take a minute to thank my parents for more things than anyone cares to read about. Words are insufficient, so I will only say here that I love you both very much and none of this would have been possible without you.

Finally, this work would not have been possible without funds from the Canadian Institute for Health Research (CIHR) and the University of Toronto.
# Table of Contents

Acknowledgments........................................................................................................ iv

Table of Contents......................................................................................................... vi

List of Tables ................................................................................................................ xi

List of Figures ................................................................................................................ xii

List of Abbreviations ................................................................................................... xiii

Chapter 1 Introduction ................................................................................................. 1

Chapter 2 Literature Review ......................................................................................... 3

1 Osteoarthritis ............................................................................................................. 3
   1.1 Definition ............................................................................................................ 3

   1.2 Epidemiology .................................................................................................... 3
      1.2.1 Prevalence .................................................................................................. 3
      1.2.2 Burden of OA ......................................................................................... 4
      1.2.3 Risk Factors ............................................................................................ 5

   1.3 Medical Management of OA .......................................................................... 6

2 Total Hip Replacement ............................................................................................ 7

   2.1 Definition ........................................................................................................ 7

   2.2 Trends in THR ................................................................................................ 8

   2.3 Outcomes of THR ......................................................................................... 10
      2.3.1 Short-term Outcomes of THR ................................................................. 10
      2.3.2 Long-term Outcomes of THR ............................................................... 10

3 Total Hip Replacement Failure ............................................................................. 11

   3.1 Definition of THR Failure ............................................................................ 11

   3.2 Predictors of THR Failure ............................................................................... 12
      3.2.1 Patient Factors ....................................................................................... 12
Chapter 5

12.5.3 The Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP) .............. 50
12.5.4 The Six Minute Walk Test (6MWT) ................................................................... 51
12.5.5 Six Minute Walk Test Pain ............................................................................... 51
12.6 Potential Covariates .............................................................................................. 51
12.6.1 Self-reported Co-morbidity Index ..................................................................... 52
12.6.2 The Pain Catastrophizing Scale ....................................................................... 52
12.7 Assessment of THR Wear .................................................................................... 52
12.7.1 The Martell Method of Wear Analysis ............................................................... 52
12.7.2 Acquisition and Preparation of x-rays ............................................................... 55
12.7.3 Evaluation of Periprosthetic Boney Change ...................................................... 55
12.8 Statistical Analysis ................................................................................................ 58
12.9 Descriptive Statistics ............................................................................................. 58
12.10 Testing the Hypotheses ....................................................................................... 58
12.11 Sample Size Justification .................................................................................... 60
12.11.1 Logistic Regression .......................................................................................... 60
12.11.2 Area under the Receiver Operating Characteristic Curve ......................... 60

Chapter 5 Results ........................................................................................................... 62

13 Results ...................................................................................................................... 62
13.1 Cohort and Sample ............................................................................................... 62
13.2 Study Sample ......................................................................................................... 62
13.3 Explanatory variables ............................................................................................ 64
13.3.1 Measures of Self-Reported Pain Post THR ....................................................... 64
13.3.2 Self-Reported Physical Function after THR ..................................................... 67
13.3.3 Walking Performance after THR ...................................................................... 68
13.3.4 Correlations of Explanatory Variables ............................................................. 69
13.4 Outcomes ............................................................................................................... 73
List of Tables

Table I: Search terms used to retrieve citations from all bibliographic databases....................... 23
Table II: Eligibility criteria for selection of title and abstracts.................................................. 25
Table III: Results of systematic literature review........................................................................ 31
Table IV: Cohort versus study sample 4-8 years after total hip replacement surgery............... 63
Table V: Characteristics of study sample 4-8 years after total hip replacement surgery........ 66
Table VI: Correlation coefficients (Pearson or Spearman Rank) for explanatory variables........ 71
Table VII: Number of participants in analysis subsample by annual wear threshold.................. 73
Table VIII: Radiographic markers of potential THR failure in study sample 4-8 years after total hip replacement surgery................................................................. 74
Table IX: Cross-tabular frequencies and agreement statistics between potential radiographic markers of total hip replacement failure............................................................. 75
Table X: Explanatory variables by wear group............................................................................. 78
Table XI: Univariate logistic regression (dependant variable = wear ≥ 0.1mm/year)............... 79
Table XII: Univariate logistic regression (dependant variable = periprosthetic boney change).. 83
List of Figures

Figure 1: Different types of right total hip replacements......................................................... 9
Figure 2: Examples of periprosthetic boney change................................................................. 19
Figure 3: Determination of 2-dimensional linear wear using the Martell method .................... 54
Figure 4: Evaluating periprosthetic boney change................................................................. 57
Figure 5: Flow diagram of cohort exclusions to get sample...................................................... 61
Figure 6: Histograms of self-reported pain scores................................................................. 65
Figure 7: Histogram of self-reported functioncores.................................................................. 67
Figure 8: Histograms of walking performance scores............................................................. 68
Figure 9: Charts of self-reported pain against annual linear wear.......................................... 77
Figure 10: Chart of self-reported function against annual linear wear..................................... 80
Figure 11: Chart of walking performance against annual linear wear..................................... 81
List of Abbreviations

6MWT: Six Minute Walk Test .................................................................................................................. 41
AP1: Unilateral anterior posterior x-ray .................................................................................................... 49
AP2: Bilateral anterior posterior pelvic x-ray ........................................................................................... 49
GDP: Gross domestic product ................................................................................................................ 5
HAS: Hip Analysis Suite .......................................................................................................................... 55
HC: Holland Orthopaedic and Arthritic Institute ..................................................................................... 48
HO: Heterotropic Ossification .................................................................................................................. 18
ICOAP: The Measure for Intermittent and Constant Osteoarthritis Pain ............................................ 43
LAT: Lateral x-ray image .......................................................................................................................... 49
NRS: Numeric rating scale for arthritis pain ........................................................................................... 52
NSAID: Non steroidal anti-inflammatory ............................................................................................... 7
OA: Osteoarthritis ...................................................................................................................................... 1
PPBC: Periprosthetic boney change ......................................................................................................... 13
SB: Sunnybrook Health Sciences Center ............................................................................................... 48
SF-36: Medical Outcome Survey Short Form ......................................................................................... 10
THR: Total hip replacement .................................................................................................................... 1
WOMAC: Western Ontario and McMaster Osteoarthritis Index ........................................................... 10, 51
WOMAC-G: WOMAC total score ............................................................................................................ 51
WOMAC-P: WOMAC pain subscale ......................................................................................................... 51
WOMAC-PF: WOMAC physical function subscale .................................................................................... 51
XLP: Highly cross-linked ultra high molecular weight polyethylene ................................................... 37
Chapter 1 Introduction

Osteoarthritis (OA) is a progressive disease of synovial joints that eventually leads to the breakdown of cartilage and bone, which manifests clinically as worsening pain, stiffness and movement restriction, crepitus, occasional effusion, variable degrees of local inflammation and eventual loss of function (Brandt, Dieppe, & Radin, 2008). OA increases in prevalence with age and is responsible for the majority of the estimated 6 billion dollars (1998 Canadian dollars) in annual Canadian health care spending on arthritis (P. C. Coyte, Asche, Croxford, & Chan, 1998a; P. C. Coyte, Asche, Croxford, & Chan, 1998b; Stokes, Desjardins, & Perruccio, 2003). With the aging Canadian population, OA is expected to be the 4th leading cause of disability by the year 2020 (Woolf & Pfleger, 2003).

The hip, one of the primary weight bearing joints of the lower extremities, is commonly affected by OA. When medical treatment no longer controls the pain, or functioning becomes unacceptable, total joint replacement of the hip (THR) is an effective means of relieving pain, and restoring function (G. A. Hawker et al., 2008). Unfortunately, prostheses can fail and, without early intervention, there may be serious consequences such as severe bone loss, hip dislocation, and fracture requiring more complex surgery with greater risk of morbidity and mortality (R. Barrack et al., 2006). Therefore, regular life-long monitoring of hip replacement patients is recommended.

Current methods of periodic post-operative monitoring involve a patient visit to the orthopaedic specialist’s office, an examination and a radiographic evaluation. In the vast majority of cases these visits identify no pathology, implying an essentially wasted visit with the associated monetary and health-human resource costs. Moreover, research shows that more than 50% of individuals who have received hip joint replacements are getting no surveillance (De Pablo et al., 2006). There are many reasons for this lack of monitoring, but among them is access to orthopaedic care. With the ever increasing prevalence of total hip replacements in Canada without a corresponding increase in trained orthopaedic health professionals, this model of surveillance is costly and unsustainable in the long term. Therefore, alternative monitoring approaches are needed.

One potential alternative surveillance strategy is to use patient self-report questionnaires, which could evaluate, for example, pain, function or a simple performance measure, in order to
identify that group of patients, at higher risk for THR failure, who require follow-up with an orthopaedic health care professional. In order to use pain, function or performance for this purpose, there is a need for research demonstrating a consistent, significant association of these patient-centered measures with radiographic markers of potential THR failure, an association which to date has not been established.

The purpose of this dissertation is, in the context of work to date, to comprehensively evaluate the cross-sectional associations of several measures of self-reported hip pain (including newly developed measures), self-reported function and walking performance with newer measures of radiographic wear and other early radiographic markers of future THR failure, in a single sample of THR recipients diagnosed with OA.
Chapter 2 Literature Review

1 Osteoarthritis

1.1 Definition

Osteoarthritis (OA) describes a potentially etiologically distinct group of diseases of the synovial joint with similar biologic, morphologic and clinical presentation resulting from a failed joint repair mechanism due to abnormal stresses on the joint. The pathophysiology of OA is multi-factorial; current research implicates complex interactions of both genetic and environmental factors causing damage to all tissue in the diarthroidal joint (Brandt et al., 2008; Goldring & Goldring, 2006). OA is not fatal; however, it is a leading cause of long term disability in the elderly, particularly women. OA is among the top 10 predictors of health problems worldwide, ranking 5th in women and 9th in men (The global burden of disease: 2004 update, 2008). It remains a priority in the World Health Organization’s proclaimed Decade of Bone and Joint Disease, emphasizing its global impact (Brooks, 2002).

1.2 Epidemiology

1.2.1 Prevalence

OA is a disease that affects the majority of Canadians by the age of 70 years (Murphy, Spence, McIntosh, & Connor, 2006). Factors affecting the prevalence of OA include age, sex, body mass index (which is controversial and supported for symptomatic definitions of OA only (S. M. A. Bierma-Zeinstra & Koes, 2007; A. M. Lieverse et al., 2002; A. M. Lieverse, Reijman, Pols, Bierma-Zeinstra, & Gelber, 2003)), region, ethnicity and occupation (Badley, 1995; Badley & Wang, 1998; D'Ambrosia, 2005; Felson & Zhang, 1998; Felson, 1998; Felson, 2004). Worldwide estimates of OA prevalence range widely from <1% in younger samples of various ethnic groups, such as the Chinese (Lau, Symmons, & Croft, 1996) and African blacks (Mijiyawa & Ekoue, 1993; Mijiyawa, 1993), to up to 27% in black and white adults age 45 and older living in the predominantly Caucasian developed world (Dagenais, Garbedian, & Wai, 2009).

The prevalence of OA increases with age, but the manner in which OA is defined for the purposes of research (on the basis of clinical symptoms, x-ray alone or a combination of both)
greatly affects the prevalence estimates. Notably, studies using radiographic definitions alone frequently show greater prevalence estimates as the relationship between structural changes and symptomatic disease in OA is inexact. Generally, symptomatic OA is accompanied by radiographic changes, although the reverse is not necessarily true. The effect of sex on prevalence estimates may vary based on the definition employed. Radiographic signs of OA tend to be more prevalent in men as men have higher rates of asymptomatic cartilaginous change (Croft, 1990; Croft, Cooper, Wickham, & Coggon, 1990; Ling & Bathon, 1998; Roux et al., 2008), while post menopausal women are as much as twice as likely to develop (Meisler, 2002) or have clinically defined OA compared to men (Rossignol et al., 2005).

In Canada, OA affects 10% of the population (Arthritis and related conditions in Ontario: ICES research atlas 2004) and its prevalence is increasing at a rate of 1.2% per year (Kopec et al., 2008). OA can affect all joints, but it is more common in the weight bearing joints of the lower extremities, such as the hip. Hip OA can be identified using either symptomatic or radiographic criteria or a combination thereof. Estimates of hip OA based on both symptomatic and radiographic data from the US are about one third the estimates (7.4% to 16%) based on radiographs alone (Croft et al., 1990; Lawrence et al., 2008; Mannoni et al., 2003; Quintana, Arostegui, Escobar, Azkarate, & Goenaga, 2008).

1.2.2 Burden of OA

Osteoarthritis is the 6th leading cause of non-fatal disability worldwide, accounting for 3% of total years lost due to disability in 2000 (The global burden of disease: 2004 update, 2008; Woolf & Pfleger, 2003). Its effect is seen on both the individual and society. Individually, 80% of people with the condition have some degree of limitation of movement and 25% cannot perform their major activities of daily life (The global burden of disease: 2004 update, 2008; Brooks, 2002). Individuals affected by OA report pain, activity limitation and a decrease in health-related quality of life (S. M. Bierma-Zeinstra & Koes, 2007; Dagenais, Garbedian, & Wai, 2009; G. A. Hawker, Wright, Coyte, Williams, Harvey, Glazier, Wilkins, & Badley, 2001; G. A. Hawker, Wright, Glazier, Coyte, Harvey, Williams, & Badley, 2002). Although only a subset of OA sufferers are employed, work-related disability in this population ranges between 30-50% (S. M. Bierma-Zeinstra & Koes, 2007; G. A. Hawker, 2009). OA sufferers are more likely to report reduced work hours, unemployment due to illness, or early retirement (Bijlsma & Knahr, 2007;
Gender impacts disability and thus the level of personal burden; it has been shown that women with OA are significantly more disabled when compared to men with OA (G. A. Hawker et al., 2000; G. A. Hawker, Wright, Coyte, Williams, Harvey, Glazier, Wilkins, & Badley, 2001; G. A. Hawker, Wright, Glazier, Coyte, Harvey, Williams, & Badley, 2002). Approximately 40% of OA patients in rheumatologic care report substantial fatigue (Power, Badley, French, Wall, & Hawker, 2008), which has elsewhere (G. A. Hawker et al., 2008; G. A. Hawker, 2009) been noted to associate with depressed mood and sleep disturbance. While the exact contribution of these findings to burden remains to be determined, all are potential contributors to the decreased health-related quality of life noted in this population (Salaffi, Carotti, Stancati, & Grassi, 2005).

From a societal perspective, OA is the most costly musculoskeletal disease worldwide (Reginster & Khaltaev, 2002) with reported estimates as high as 2.5% of a country’s gross domestic product (GDP) (Bitton, 2009; The global burden of disease: 2004 update, 2008; Woolf & Pfleger, 2003). As OA is the most prevalent form of arthritis and rheumatism, it accounts for the majority of costs for these conditions, costs which are reported to be between 4.4 and 6.6 billion in 1998 Canadian dollars (Arthritis and related conditions in Ontario: ICES research atlas, 2004; P. C. Coyte, Asche, Coxford, & Chang, 1998; P. C. Coyte et al., 1998; Stokes et al., 2003). While substantial, this figure likely underestimates the true costs, as studies in Ontario suggest that the true societal costs are almost 20 times the annual per person estimate used in the cost estimates above ($12,200 in 2002 Canadian dollars versus $700 in 1994 Canadian dollars) if one includes indirect costs such as the value of patient time lost and the value provided by unpaid caregivers (Gupta, Hawker, Laporte, Croxford, & Coyte, 2005). Predictors of increased disease-related expenditures include female gender, age extremes, severe disease status (Bijlsma & Knahr, 2007; March & Bachmeier, 1997), and having co-morbidities (Scher, Belmont, P.J, Mountcastle, & Owens, 2009).

1.2.3 Risk Factors

1.2.3.1 Incidence

While it is evident that OA is prevalent and poses a significant individual and societal burden as noted above, there are certain factors that increase an individual’s chances of
developing OA. They include: advanced age, female sex, obesity, abnormal joint mechanics or anatomy, genetics, bone density, nutritional factors and reproductive hormones (Sowers, 2001). Recently, occupational factors (Axmacher & Lindberg, 1993; A. Lievense, Bierma-Zeinstra, Verhagen, Verhaar, & Koes, 2001; Rossignol et al., 2005), muscle weakness, physical activity, injury (joint damage), and intensive sports participation were added to the growing list (S. M. Bierma-Zeinstra & Koes, 2007; A. M. Lievense, Bierma-Zeinstra, Verhagen, Verhaar, & Koes, 2002).

1.2.3.2 Progression

Progression of hip OA generally results in increased pain, decreased function and decline in quality of life. Deterioration in patients with hip OA can be rapid, with a significant proportion of sufferers progressing beyond medical management to joint replacement within 5 years (Osteoarthritis: National clinical guideline for care and management in adults. 2008; A. M. Lievense, Koes, Verhaar, Bohnen, & Bierma-Zeinstra, 2007). Risk factors for the progression of OA identified in systematic reviews include age at diagnosis and certain structural changes within the joint. Specifically, older age at entry into the health care pathway with OA related issues, superior-lateral migration of the femoral head, and atrophic bone response are strong predictors for the progression of radiographic and clinical hip OA (S. M. Bierma-Zeinstra & Koes, 2007; A. A. Wright, Cook, & Abbott, 2009), while joint space narrowing, a frequently used indicator of disease progression, was recently noted to have limited (clinical OA) and inconsistent (radiographic OA) support as an indicator of OA progression to total hip replacement surgery (S. M. Bierma-Zeinstra & Koes, 2007). These limitations are largely due to various methods of use (alone or within a composite measures) and thresholds for what constitutes joint space narrowing (< 1.0 mm - < 2.5 mm) (A. A. Wright et al., 2009) which make it difficult to evaluate the effect of joint space narrowing between studies or independently of other measures of radiographic change. Composite measures like the Kellgren-Lawrence scale (≥ 3) which includes joint space narrowing is generally considered a strong predictor of OA progression to total hip replacement surgery (A. A. Wright, Cook, & Abbott, 2009).

1.3 Medical Management of OA

Medical management of OA is not curative, but focuses on treating pain and stiffness, maximizing joint function, preventing further deterioration, and improving health-related quality

Recommended first line oral analgesics include, initially, acetaminophen, followed by non-selective non-steroidal anti-inflammatory (NSAID) agents. Second line therapy such as intra-articular injections of corticosteroids (for pain management) or hyaluronate (as a joint lubricant) may also be considered (Zhang et al., 2007). When these choices are no longer effective or contraindicated, third line treatment with the lowest effective dose of opioid analgesics is suggested (Chang, Pellisier, & Hazen, 1996; G. A. Hawker, Wright, Badley, & Coyte, 2004). When medical management fails, total hip replacement is the recommended treatment (G. A. Hawker et al., 2000; G. A. Hawker, Wright, Coyte, Williams, Harvey, Glazier, Wilkins, & Badley, 2001; G. A. Hawker, Jamal, Ridout, & Chase, 2002; Hudak et al., 2002).

2 Total Hip Replacement

2.1 Definition

Total hip replacement (THR) is a surgical procedure, which is exceptionally effective at relieving pain and restoring function in patients with symptomatic OA of the hip (Arthritis and related conditions in Ontario: ICES research atlas, 2004; Brooks, 2006; Canadian Joint Replacement Registry (CJRR) 2007 Annual Report—Hip and Knee Replacements in Canada (Ottawa: CIHI, 2008)). A total hip replacement (arthroplasty) involves replacing both the hip socket (the acetabulum) and the femoral head with artificial components, which are fixed into the thigh bone with a metal post (the femoral stem) as shown in Figure 1. The artificial components can be secured to the underlying bone with cement (“cemented total hip replacement”). Alternatively, specialized metal components have a surface that allows for bone ongrowth or
ingrowth, thus securing the components without the use of cement (“uncemented total hip replacement”). When cement is used on only one side (usually the femoral side), the term “hybrid total hip replacement” is used. In a traditional hip replacement, a metal shell sits in the hip socket with an ultra high molecular weight ‘polyethylene liner’ inside to mimic the joint cartilage and make up the bearing surface of the joint. More modern alternatives use either a highly cross-linked polyethylene, metal or ceramic liner, or avoid the metal shell entirely and cement a single component entirely made of ultra high molecular weight polyethylene directly to the hip socket. A metal or ceramic head attached to the femoral stem forms the other half of the bearing surface (Hip replacement: Current trends and controversies 2002; R. Barrack et al., 2006).

2.2 Trends in THR

The number of joint replacements per year is increasing globally. In Canada during 2005-2006, 76.5 primary THRs were performed per 100,000 population. There were 67.8 and 88.6 THR surgeries per 100,000 population in men and women, respectively, compared to 47.4 in men and 65.9 women in the previous decade. This represents a minimum 35% increase in the number of THRs performed (Canadian Joint Replacement Registry (CJRR) 2007 Annual Report—Hip and Knee Replacements in Canada (Ottawa: CIHI, 2008)). Explanations for this increase include an ageing population with additional and more severe cases of OA (Arthritis and related conditions in Ontario: ICES Research Atlas, 2004), more inclusive eligibility criteria (Access to health services in Ontario: ICES atlas, 2006), and an increased willingness on the part of younger patients to consider THR (G. A. Hawker et al., 2004; G. A. Hawker, 2006).
**Figure 1:** Different types of right total hip replacements.  A) Hybrid: Metal on polyethylene bearing surface.  B) Cemented: All polyethylene cup.  C) Uncemented: Ceramic head on polyethylene bearing surface.  D) Uncemented: Metal on metal bearing surface.
2.3 Outcomes of THR

2.3.1 Short-term Outcomes of THR

THR is among the most effective of medical interventions. It relieves pain, restores function and improves health-related quality of life (Hip and Knee Replacements in Canada—Canadian Joint Replacement Registry (CJRR) 2008–2009 annual report). The majority (> 90%) (Fortin et al., 1999) of patients are mobile within days to weeks of the surgery and report substantial if not complete relief of pain using a variety of commonly used OA measures (Jones, Beaufre, Johnston, & Suarez-Almazor, 2007). Large improvements in function have been consistently reported in the literature, as measured by the Western Ontario and McMaster Osteoarthritis Index (WOMAC) and the Medical Outcome Survey Short Form (SF-36) (Jones, Voaklander, Johnston, & Suarez-Almazor, 2000; Canadian Joint Replacement Registry (CJRR) 2007 Annual Report—Hip and Knee Replacements in Canada (Ottawa: CIHI, 2008)). Improvements in health-related quality of life and walking performance, measured by the six minute walk test, have been shown as early as three months post-surgery (Jones et al., 2000; Laupacis et al., 1993).

Surgical complications (rates in the United States unless otherwise specified) do occur and include: infection (1%), periprosthetic fracture (0.3% cemented and 5.6% uncemented), hip dislocation (3%), leg-length discrepancy (up to 50% with 15-20% requiring contralateral shoe correction) (R. Barrack et al., 2006), vascular injury (rare <1%) (Abularrage et al., 2008), nerve injury (rare 0-3%; 0.03% (Yacub, Rice, & Dillingham, 2009)), and significant blood loss requiring blood transfusion. However, figures show a decrease in the relative percentages of complications in recent years (R. Barrack et al., 2006) supporting the claim that, in the short-term, the results of THR are generally excellent and improving.

2.3.2 Long-term Outcomes of THR

Long-term survival rates for THR prostheses are improving with advancing implant technology and surgical technique. Survival rates of approximately 90% at 10 years for implants manufactured in the 1970’s and early 1980’s have improved to 90-100% at up to 20 years for implants manufactured in the mid 1980’s and later (Aldinger et al., 2009; Merle, Clarius, & Aldinger, 2009; Müller et al., 2009; Morshed, Bozic, Ries, Malchau, & Colford, 2007).
Survival rates vary by implant but meta-analysis has confirmed they are (K. T. Makela, Eskelinen, Pulkkinen, Paavolainen, & Remes, 2008) similar for cemented and uncemented prosthetics. Makela et al. examined the twelve most common cemented implants used in the Finnish Joint Replacement Registry between 1980 and 2005 in patients age 55 and older and confirmed the previously noted trends; they reported survival rates between 70-90% at 10 years (K. Makela, Eskelinen, Pulkkinen, Paavolainen, & Remes, 2008). North American results for uncemented designs (more common in Canada and the United States), with some notable exceptions (Dobzyniak, Fehring, & Odum, 2006), have shown similar results with 10-20 year survival rates reported to be above 80%, and in most cases greater than 90% (Aldinger, Thomsen, Mau, Ewerbeck, & Breusch, 2003; Aldinger et al., 2009; Huo, Gilbert, & Parvizi, 2007; Merle et al., 2009; Yoon, Rowe, Kim, Cho, & Seon, 2008). Equivalent survival is particularly noteworthy as uncemented implants are generally used in younger (generally more active) patients who traditionally have poorer long term outcomes (Aldinger et al., 2009; Flecher, Pearce, Parratte, Aubaniac, & Argenson, 2009; Johnsen, Sorensen, Pedersen, Lucht, Soballe, & Overgaard, 2006; Keener et al., 2003). These dramatic improvements in long-term survival are generally attributed to a better understanding of the causes of failure, such as polyethylene bearing surface wear, and how to mitigate their effects. Despite the improved understanding, the risk of failure increases with time elapsed after THR surgery. Given the increases in life expectancy, there is the possibility that the THR will eventually fail during the patient’s lifetime.

3 Total Hip Replacement Failure

3.1 Definition of THR Failure

THR failure is defined as the requirement for a second or revision THR on the same joint, where the hardware of the acetabulum or the femoral component, or both, is replaced. Early identification of and intervention in patients at risk for impending THR failure is crucial to achieving the best patient-centered and economic outcomes as delays are associated with complex revision surgery with increased morbidity, mortality and cost (A. M. Davis et al., 2006; A. M. Davis et al., 2008).
3.2 Predictors of THR Failure

There is an extensive body of literature examining predictors of THR failure. Predictors of revision THR surgery can be subdivided into three broad categories: patient factors, surgeon factors and implant factors.

3.2.1 Patient Factors

A 2008 systematic review examined patient-centered factors related to failure and identified 38, primarily retrospective cohort studies, with sample sizes over 500 and with extractable data for patients with OA (Santaguida et al., 2008). Increased risk for revision THR surgery (variably defined) included patients younger at surgery (Aldinger et al., 2009; Flecher, Pearce, Parratte, Aubaniac, & Argenson, 2009; Johnsen, Sorensen, Pedersen, Lucht, Soballe, & Overgaard, 2006; Keener et al., 2003) and male sex. A nationwide Danish follow-up study based on Danish Hip Registry data supported male sex as a predictor and identified a high Charlson co-morbidity index score (high scores (>2) with adjusted relative risk of 2.8 (2.3 to 3.3) compared to low scores (0)) as predictive of failure. Overall, patient factors included age <40 (Johnsen, Sorensen, Pedersen, Lucht, Soballe, & Overgaard, 2006), longer time since surgery, high activity level, lack of appropriate follow-up (Hulleberg, Aamodt, Espehaug, & Benum, 2008; Needham, Burns, & Gerlinger, 2008), male sex (3 to 5 fold increase compared to women, largely due to aseptic loosening) (Santaguida et al., 2008), underlying diagnosis (congenital diseases, avascular necrosis), patient bone quality (Nixon, Taylor, Sheldon, Iqbal, & Harper, 2007), certain occupations (heavy manual labour such as construction or stone masonry compared to office or clerical positions), and greater co-morbidity (Crawford & Murray, 1997).

The cause of the increased risk of THR failure in young patients is unknown but suggested factors include the different characteristics of this group (Flecher et al., 2009). A recent meta-analysis by Jager et al. (Jager et al., 2008) noted that conclusions regarding younger patients are difficult to make with the heterogeneous populations that are being considered. Nevertheless, it is clear that younger patients tend to be more active than their older counterparts, undergo THR for structural or systemic deformities, and have fewer co-morbidities. They are consequently predisposed to increased polyethylene wear and a variable disease prognosis (Jager et al., 2008) which may consequently elevate their risk for early THR failure.
3.2.2 Surgeon Factors

Surgeon factors are generally related to experience and choice of surgical approach. Low-volume surgeons have worse outcomes (Bordini et al., 2007; Losina et al., 2004; Losina, Barrett, Mahomed, Baron, & Katz, 2004). Poor component positioning and a postero-lateral surgical approach (Arthursson, Furnes, Espenhaug, Havelin, & Soreide, 2007) have also been identified in prospective cohort studies as surgeon-controlled factors that increase the risk of prosthesis failure (Witzleb, Stephan, Krummenauer, Neuke, & Gunther, 2009). A 2006 Cochrane review (Jolles & Bogoch, 2006) concluded that there was insufficient evidence to support an ideal approach for THR for OA.

3.2.3 Implant Factors

Finally, implant factors such as implant design (stem geometry/material/coating), fixation method, and material processing (gamma irradiation in air for polyethylene liners) have been shown to have a significant effect on THR survival (Hip replacement: Current trends and controversies, 2002; R. Barrack et al., 2006; R. Barrack et al., 2006), although the strength of this effect is declining with improved implant design (Broos & Fourneau, 2000). Polyethylene wear, periprosthetic boney change (PPBC), mechanical (a.k.a. aseptic) loosening, fracture of the implant or surrounding bone, dislocation and/or instability, component mal-positioning, impingement (catching of the acetabular rim on the proximal femur), and infection (sepsis) may all increase risk for THR failure (Bozic et al., 2009). Of these, polyethylene wear and periprosthetic boney change (osteolysis, osteopenia, cortical hypertrophy, pedestal, heterotopic ossification) have been identified as both directly and indirectly responsible for the majority of revision surgeries in North America (Marshall et al., 2008; Hip and Knee Replacements in Canada—Canadian Joint Replacement Registry (CJRR) 2008–2009 annual report).

4 Radiographic Markers of Poor Long-Term Outcome or Potential THR Failure

4.1 Polyethylene Wear

Polyethylene wear is frequently the cause of bone destruction, such as pelvic osteolysis, culminating in component failure. It is detectable early on and thus a useful early marker for assessing patient risk of THR failure (Marshall, Ries, & Paprosky, 2008).
4.1.1 Definition

Wear and/or polyethylene wear are terms that are generally used interchangeably in the THR literature when discussing metal or ceramic on polyethylene bearing surfaces. Wear in materials science is “the erosion of material from a solid surface through interaction with another surface”, which generally occurs as a result of mechanical action (Rabinowicz, 1995). Polyethylene wear refers to the processes that generate polyethylene particles in addition to the normal adaptation (permanent deformation of the liner without particle release) of the new implant in vivo that occurs upon initial weight bearing.

4.1.2 Mechanisms

Mechanisms of wear include adhesion, abrasion and fatigue of the material leading to the release of material particles, called polyethylene wear debris, from the surface of the implant. By far the most common source of polyethylene wear debris is from articulation at the bearing surface, which is required for a well-functioning hip implant. It is suggested that these particles are causative agents driving a series of biological responses causing the hip to loosen, become painful and fail (Aspenberg & Van der Vis, 1998; Kobayashi et al., 1997). Current estimation methods involve evaluating the mean linear penetration rate (described below) and from it, calculating volumetric wear. The number of particles is then estimated based on the average volume of polyethylene lost.

4.1.3 Predictors of Wear

Patient variables that are linked to excessive wear include level of physical activity and age. BMI may also affect function (Santaguida et al., 2008), and thus, the rate of wear (Gallo, Havranek, & Zapletalova, 2010).

Since polyethylene wear is related to use (load cycles), more active younger patients will accumulate more load cycles over a given time period and generate, as a result, increased wear. Younger patients may also engage in different physical activity profiles than their older counterparts, activities that may employ different path motions within the joint which have been linked to higher wear. The interplay and relative impact of age (Johnsen, Sorensen, Lucht et al., 2006), and physical activity on polyethylene wear is likely complex and poorly understood (Munger,
Roder, Ackermann-Liebrich, & Busato, 2006; Rosenbaum, Bloebaum, Ashrafi, & Lester, 2006; Bennett et al., 2008).

The influence of BMI on wear is also unclear. While a 2007 Norwegian Registry study showed an increased relative risk of failure, defined as aseptic loosening in the stem or both components, for men and women with BMIs above the seventy-fifth percentile, relative to their normal weight counter-parts (BMI < 24.3 kg/m²) (Flugsrud, Nordsletten, Espehaug, Havelin, & Meyer, 2007), there is a large body of literature that consistently supports no increased risk of excessive polyethylene wear, failure or poorer long-term outcome in obese individuals compared to their normal weight counter-parts (Andrew et al., 2008; Cushnaghan et al., 2007; W. Davis & Porteous, 2007; Gallo et al., 2010; Haverkamp, de Man, de Jong, van Stralen, & Marti, 2008; Kessler & Kafer, 2007; Andrew et al., 2008; Gallo et al., 2010).

4.1.4 Measuring Polyethylene Wear

Measures of evaluating polyethylene wear have improved in recent years with advancing technology; newer technologies may provide a more sensitive and specific measure of “risk for failure” than has previously been available. In 2005, McCalden and colleagues reviewed 11 of the methods currently in use to evaluate wear. Overall conclusions indicated that computer techniques resulted in improved precision (0.072 versus 0.62 mm) and accuracy (0.033 versus 1.64 mm) compared to older manual techniques (such as the Livermore method) which are done by hand and/or by using digital callipers (McCalden, Naudie, Yuan, & Bourne, 2005). Two computer-assisted techniques based on edge detection comparison of serial radiographs commonly used for research are: The Martell Hip Analysis Suite and Devane’s PolyWare program. Each has strengths and limitations (Devane, Horne, Martin, Coldham, & Krause, 1997; McCalden et al., 2005), but with optimal and sub-optimal radiographs, Martell’s method was shown to be superior to Devane’s PolyWare Suit (Sychterz, Young, & Engh, 2001).

4.1.4.1 The Martell Method

Initial data using Martell two-dimensional wear analysis techniques demonstrated impressive accuracy and precision when calculating two-dimensional linear wear. In an early validation study comparing results obtained using the Martell method to known values of wear established from x-rays of constructed biomechanical pelvic models (phantoms), accuracy estimates of 0.080 ± 0.033 mm and precision values of 0.060 ± 0.072 mm were reported.
Unfortunately, these estimates relied on a small series of high quality x-rays which did not represent normal clinical study conditions (Martell & Berdia, 1997). A more recent comparison, using both high and low quality clinical and phantom (ex-vivo) images, provided a mean measurement error of 0.13 mm, much less than that of digital callipers (±0.5 mm), and other methods previously employed (Sychterz et al., 2001). Hui et al. showed a 19% absolute difference between the radiographic estimate and wear at revision retrieval, with a mean interval of 9.2 years before liner extraction, and excellent agreement ($r^2 = 0.80$) between two-dimensional Martell radiographic linear wear estimates and a coordinate measuring machine (Hui et al., 2003).

While the Martell method shows good accuracy and is applicable for use in large cohort studies (McCalden et al., 2009), due to the method’s relatively low cost (Bragdon et al., 2005) and absence of need for a prospective design, significant variation is introduced into these estimates by poor or variable image quality occasionally seen in the clinical setting (Geerdink, Grimm, Vencken, Heyligers, & Tonino, 2008; Hui et al., 2003; McCalden et al., 2005; McCalden et al., 2009). Thus, there is debate regarding the validity of the Martell technique for small measures of wear below 0.1 mm. Nevertheless, this technique has been used to demonstrate significant correlations with osteolysis at measurement thresholds below 0.2 mm/year (Dumbleton, Manley, & Edidin, 2002; Orishimo, Claus, Sychterz, & Engh, 2003).

### 4.1.5 Polyethylene Wear Thresholds for Osteolysis

Clinical investigations into the nature of wear associated with failure have noted thresholds, above which the amount of polyethylene debris is correlated with significant osteolysis or loosening, placing the joint at risk. The actual amount of wear that appears to trigger significant osteolysis is controversial (>0.1 or >0.2 mm/year), and fraught with methodological challenges (Dumbleton et al., 2002), but annual wear rates below 0.1 mm/year have consistently been associated with a lower incidence of osteolysis while increasing linear wear rates above this threshold quadruple the likelihood of osteolysis (R. Barrack et al., 2006). Although up to 10% of patients can develop osteolysis at wear rates below 0.1 mm/yr, it has been suggested that wear below this threshold drastically reduces or even eliminates (below 0.05 mm/year) osteolysis (Dumbleton et al., 2002; Harris, 2003; Holt, Murnaghan, Reilly, & Meek, 2007; Shon et al., 2009).
4.2 Periprosthetic Boney Change

Periprosthetic boney changes around the prosthesis have also been linked to poor long term outcome or THR failure (Bozic et al., 2009; Marshall et al., 2008; Hip and Knee Replacements in Canada—Canadian Joint Replacement Registry (CJRR) 2008–2009 annual report).

4.2.1 Definition

Periprosthetic boney change refers to 5 modes of bone remodelling: osteolysis, osteopenia, cortical hypertrophy, pedestal and heterotopic ossification, that occur around the prosthesis between the first post-operative and last follow-up x-ray. Osteolysis refers to bone destruction (re-absorption) which manifests radiographically as radiolucent areas around the prosthesis. Osteopenia secondary to stress shielding, which appears as a subtle increased translucence within the bone on film, refers to a decrease in bone density due to non-anatomic load distributions around the prosthesis. Cortical hypertrophy is an increase and bowing of the outer region of the bone, the cortex, in the bottom third of the femoral stem. Pedestal refers to the formation and mineralization of a bone plate across the center of the bone perpendicular to the tip of the femoral stem and heterotopic ossification describes the formation and mineralization of bone in the muscle and fibrous tissue surrounding the total hip replacement.

4.2.2 Measuring Periprosthetic Boney Change

Evaluating periprosthetic boney change is more subjective than the computerized methods for evaluating wear and is based on standardized expert evaluation of serial radiographs. Experts assess the presence or absence of significant change in the 5 modes of bone remodelling in all of three zones in the acetabulum and 14 zones around the femoral component (Figure 2). The 14 zones are those proposed by DeLee and Charnley, which were expanded by Gruen and later Johnston into the current standardized method for evaluating clinically silent boney change (DeLee & Charnley, 1976; Gruen, McNeice, & Amstutz, 1979; Johnston et al., 1990). The validity and reliability of assessing PPBC in this manner is also dependent on the training of the clinicians involved, with substantial disagreement between generalists and specialists in the field (Kocic et al., 2010; Röder et al., 2003; Toms et al., 2009). For example, the inter-rater reliability in evaluating osteolysis on x-rays is good to excellent when the evaluation is based on the
relative change between two radiographs at different time points, but is notoriously poor when based only on a single film (Kocic et al., 2010; Röder et al., 2003; Toms et al., 2009).

### 4.2.3 Periprosthetic Boney Change as a Marker of Potential THR Failure

Not all the measures of periprosthetic boney change discussed above are equally indicative of future THR failure. Severe osteolysis in the form of complete radiolucent lines around the cement mantel of the acetabular cup has a 94% sensitivity for detecting aseptic loosening of this component, and lesser extremes still show a proportional relationship with the degree of osteolysis related to the degree of radiographic loosening. Certain zones or patterns of radiolucencies corresponding to osteopenia and osteolysis are correlated with loosening and subsequent failure (Toms et al., 2009). However, implants can be radiographically loose but stable intraoperatively.

In the event of large-scale bone demineralization (>70%), stress shielding is reliably detected as osteopenia. Although significant osteopenia appears to be relatively benign, it carries an increased risk of implant instability or periprosthetic fracture (Thomsen et al., 2008; Whirlow & Rubash, 1995). Fortunately, this is rare in short-term post-operative THR recipients, making this outcome of limited value in evaluating very early radiographic change. As cortical hypertrophy and pedestal alone are also relatively rare, there is no data assessing their ability to predict failure. However, they are indicators of implant micro-motion or abnormal load distribution which may eventually worsen and result in implant instability and THR failure (Kocic, Lazovic, Mitkovic, & Djokic, 2010; Röder, Eggli, Aebi, & Busato, 2003; Toms, Botchu, & Nolan, 2009).

Heterotopic ossification (HO) is classified using Brookers grades 1-4 corresponding to the amount of uncontrolled bone growth (Schuh and Zeiler 2005). It is often asymptomatic and does not impact bone stability but can severely impact patient function and may be associated with late pain (Schuh & Zeiler, 2005; Kocic et al., 2010).

Ideally the 5 components of PPBC would be assessed separately. However, due to the rarity of PPBC at early follow-up, a composite measure may also be employed to augment statistical power for analysis.
Figure 2: Examples of periprosthetic boney change. Top row: 28 to 90 days post THR. Bottom row: 4-8 years post THR. A-B) Osteolysis. C-D) Osteopenia. E-F) Cortical hypertrophy (mild) and development of pedestal. G-H) Heterotopic ossification Brooker’s grade 3.
5 Consequences of THR Failure

THR failure, while rare, does occur and the incidence increases with the length of time since implantation. Consequences of THR failure, without early intervention, include massive host bone and tissue destruction (potentially culminating in component migration), femoral head dislocation, or periprosthetic (pelvic or femoral) fracture with their associated medical sequelae (arterial bleeding, patient injury, long-term functional limitations, mortality) (Canadian joint replacement registry (CJRR) 2007 Annual Report—Hip and Knee Replacements in Canada (Ottawa: CIHI, 2008)). The greater the destruction of bone, the longer and more complex will be the revision surgery (Rubash, Sinha, Shanbhag, & Kim, 1998). This increases patient exposure to infectious agents (Jafari, Coyle, Mortazavi, Sharkey, & Parvizi, 2010) and thus to higher patient morbidity and mortality (Paprosky, Bradford, & Jablonsky, 1995). Therefore, early identification of individuals at increased risk for THR failure and timely intervention is necessary.

6 Surveillance for THR Failure

6.1 Current Methods

In view of the potentially serious consequences of THR failure, regular lifelong monitoring of post-surgical THR patients is recommended (R. Barrack et al., 2006; Hodgkinson et al., 1988). While there are no evidence-based guidelines for the manner, methods (R. Barrack et al., 2006; Stulberg, Wixson, Adams, Hendrix, & Bernfield, 2002) or frequency of post-operative surveillance (Teeny, York, Mesko, & Rea, 2003), recommendations out of the Implant Wear Symposium 2007 Clinical Work Group advise radiographic follow-up at 1, 5 and 10 years after surgery, and every one to five years thereafter, depending on the radiographic findings of osteolysis and its progression (Malchau, Potter, & Implant Wear Symposium 2007 Clinical Work Group, 2008). In Ontario, existing approaches to monitoring for prosthesis failure usually entail 2-3 follow-up appointments in the first year, followed by anywhere from no further evaluation to regular x-rays and visits with an orthopaedist in the outpatient setting (Ontario data - unpublished).
### 6.2 Non-Sustainability of Current Methods

Unfortunately, the cost (Lee, Lim, Kong, & DeLisa, 2001) and health-human resource requirements for long-term, in-person surveillance is high (Bolz, Crawford, Donnelly, Whitehouse, & Graves, 2010). Even with a significant proportion of patients (>60%) failing to comply with any recommended monitoring after the first post-operative year, it is a challenge to the over-burdened health care system to provide those returning patients with biennial or regular in-person follow-up (Clohisy, Kamath, Byrd, Steger-May, & Wright, 2008; De Pablo et al., 2006). For example, while the number of THR procedures (set to double by 2020 (Canadian Joint Replacement Registry (CJRR) 2007 Annual Report—Hip and Knee Replacements in Canada (Ottawa: CIHI, 2008)) and THR survival are increasing, the number of orthopaedic clinical specialists trained annually is not increasing at the same rate (Comeau, 2004). As surgeons currently spend 2/3 of their time outside the operating room, with a significant portion of this time allocated to outpatient follow-up visits, it is clear that, moving forward, there will be an increasing deficit of health-human resources in relation to the projected demand. In order to alleviate the pressure on an already overburdened health care system, alternative methods of surveillance, that could be employed at home or in the primary care setting, are needed in order to provide adequate and accessible patient monitoring, minimize health care costs, and reallocate surgeon time to the operating room where it is most needed (Shipton, Badley, & Mahomed, 2003).

### 6.3 Alternative Methods of Surveillance

Successful alternative strategies should minimize health care costs, provide reliable and accessible monitoring of patients, and target surgeon follow-up to patients most likely to need specialist intervention. To meet these needs, methods that can be employed at home or in the primary care setting have been proposed (Ghoz & Macdonald, 2008). One potential alternative strategy for surveillance would be the use of measures of joint pain and functional limitations to identify patients at risk for future THR failure. Many measures exist to assess pain and functional difficulties and objective performance in a post-THR population. As pain and functional deficits are the most common patient complaints when patients present to the orthopaedic surgeon with a failing hip (R. Barrack et al., 2006), the possibility of exploiting these constructs for the purposes of surveillance should be considered. A recent focus on the importance of an objective measure of function (a performance measure), such as a walk test, suggests that it
should also be considered (Stratford & Kennedy, 2006), since walking performance tests have been shown to be complementary to self-reported function, which may only assess a patient’s perception of the experience.

In other parts of the world, such alternative measures are already being considered. Ghoz et al. endorsed follow-up based on pain or functional questionnaires in Britain, provided that surgeons using these methods were only prepared to revise a ‘failed’ THR if the patient has some pain or functional impairment (Ghoz & Macdonald, 2008). However, such practices are controversial, and clinical and radiographic follow-up is consistently recommended (Primary total hip replacement: A guide to good practice, 2006; Bolz et al., 2010; Teeny, York, Mesko, & Rea, 2003). In order to assess patient self-report or performance measures as potential surveillance tools, or to even consider testing their ability to discriminate patients ‘at-risk’ for failure, demonstration of a consistent association with THR failure itself, or radiographic markers of potential THR failure, such as radiographic wear or PPBC, is necessary.

7 Surveillance using Pain, Function or Performance Measures

Employing a screening tool to identify ‘at-risk’ patients with significant polyethylene wear or periprosthetic boney change before scheduling an in-person specialist evaluation would be an effective means of providing regular, cost-effective, accessible, post-operative surveillance with minimal patient and surgeon burden. Pain and function and walking performance are common patient-centered constructs utilized in the OA population that have been proposed as candidate measures for such a screening tool. A systematic search of the literature was performed to assess the existing evidence relating self-reported pain or function, or walking performance to THR failure (defined variably as any of: revision surgery on any component, or evidence of radiographic markers such as polyethylene wear or measures of periprosthetic boney change).
7.1 Search Strategy

A systematic review of 7 databases (COCHRANE LIBRARY (Inception to Feb 2010), PUBMED (1865 to Feb 2010), MEDLINE (1950 to February Week 1 2010), EMBASE (1980 to 2010 Week 06), AMED (1985 to February 2010), HEALTHSTAR (1966 to January 2010) and CINAL (1981 to February 15 2010)) was performed to determine the existing evidence of a relationship of self-reported pain or function or walking performance to THR failure or its precursors. A consistent search strategy, as detailed in Table I, was used where search strategy terms were identified using the PICO method (Bhandari, Guyatt, Montori, Devereaux, & Swiontkowski, 2002). The search was pilot-tested starting in July/August 2008 and was modified as needed. The search strategy was verified by two independent and experienced research librarians (McKibbon, 1999).

<table>
<thead>
<tr>
<th>Table I: Search terms used to retrieve citations from all bibliographic databases.*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 1</strong>: (note did not limit to studies of OA due to small number of available studies)</td>
</tr>
<tr>
<td>Arthroplasty, Replacement, Hip</td>
</tr>
<tr>
<td><strong>Set 2</strong></td>
</tr>
<tr>
<td>Pain+ (Function or Impairment or Disability)</td>
</tr>
<tr>
<td>Performance+</td>
</tr>
<tr>
<td>“Self-report”+</td>
</tr>
<tr>
<td>questionnaires+</td>
</tr>
<tr>
<td><strong>Set 3</strong>:</td>
</tr>
<tr>
<td>“Total Hip Replacement Failure”</td>
</tr>
<tr>
<td>Revision+</td>
</tr>
<tr>
<td>Postoperative Complications+</td>
</tr>
<tr>
<td>Equipment Failure+</td>
</tr>
<tr>
<td>Prognosis+</td>
</tr>
<tr>
<td>&quot;Outcome Assessment &quot;+</td>
</tr>
<tr>
<td>Treatment Outcome+</td>
</tr>
<tr>
<td>Reoperation+</td>
</tr>
<tr>
<td>Survival</td>
</tr>
<tr>
<td>“Heterotopic Ossification”</td>
</tr>
<tr>
<td>“Osteolysis”</td>
</tr>
<tr>
<td>Bone remodelling</td>
</tr>
<tr>
<td><strong>Set 1 and Set 2 and Set 3</strong></td>
</tr>
</tbody>
</table>

*Terms varied with the database [individual controlled vocabulary], but the same strategy was used. +The term was exploded to include subheadings (e.g. if Arthroplasty was exploded and had hip and knee as subheadings all article under each of the three would be retrieved)
7.2 Study Identification & Selection

7.2.1 Identification

A total of 9287 studies were identified using the search strategy outlined above. In order to identify relevant articles on the basis of title and abstract, all citations were imported into reference management software in a stepwise manner, with the removal of duplicates during each importation.

7.2.2 Selection

All citations were reviewed by a single rater based on an a priori set of selection criteria (Table II). Articles without abstracts in either English or French were excluded. Articles focused on alternative bearing surfaces, patient populations other than OA or those making no mention of both clinical and radiographic follow-up in the manner of interest were excluded on the basis of title or abstract. If the rater was undecided, the full text of the article was retrieved and evaluated for relevance. Twenty-seven articles were identified for complete evaluation. Upon full text review, 15 of these 27 articles were selected as having relevant and extractable data.

References from the 27 relevant articles were scanned for any missed articles (Birch et al., 2003). Additional searches were performed where necessary.
### Table II: Eligibility criteria for selection of title and abstracts.

<table>
<thead>
<tr>
<th>Population:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion:</strong></td>
</tr>
<tr>
<td>• Humans with total hip arthroplasty surgical intervention.</td>
</tr>
<tr>
<td>• Implants being considered had a metal on polyethylene bearing surface or some indication thereof.</td>
</tr>
<tr>
<td>• Minimum of 3 years post-index procedure</td>
</tr>
<tr>
<td>• At least some replacements as a result of osteoarthritis</td>
</tr>
<tr>
<td><strong>Exclusion:</strong></td>
</tr>
<tr>
<td>• Studies focused on any other surgical intervention (Osteotomies, debridement, hemi-arthroplasty etc.)</td>
</tr>
<tr>
<td>• Only alternative bearing surfaces considered (metal on metal, ceramic on ceramic or a combination).</td>
</tr>
<tr>
<td>• Population entirely with inflammatory arthritis or congenital defects with expected prognosis different than that of patients with arthroplasty due to OA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposures:</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least one of the following present with one of the outcomes listed below:</td>
</tr>
<tr>
<td><strong>Standardized clinical evaluation</strong></td>
</tr>
<tr>
<td>• Used standardized method (e.g.: Harris Hip, Oxford Hip Score, etc)</td>
</tr>
<tr>
<td><strong>Evaluation of Pain</strong></td>
</tr>
<tr>
<td>• Rating presence or absence, scale, standardized instrument.</td>
</tr>
<tr>
<td><strong>Evaluation of Function/Disability/Impairment</strong></td>
</tr>
<tr>
<td>• Rating presence or absence, scale, standardized instrument</td>
</tr>
<tr>
<td><strong>Evaluation of objective or walking performance</strong></td>
</tr>
<tr>
<td>• Evaluated using a standardized objective measure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome:</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least one of the following was present:</td>
</tr>
<tr>
<td><strong>Failure</strong></td>
</tr>
<tr>
<td>• Rigorously defined as revision surgery of any component or all components.</td>
</tr>
<tr>
<td><strong>Radiographic indicators</strong></td>
</tr>
<tr>
<td>• including any of: penetration, loosening, osteolysis, osteopenia, heterotopic ossification, cortical hypertrophy or pedestal</td>
</tr>
<tr>
<td>• Method noted, preferably established using validated criteria (e.g.: DeLee, Gruen, Johnston, Hodgkinson)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RCT</strong></td>
</tr>
<tr>
<td><strong>Cohort</strong></td>
</tr>
<tr>
<td>• Prospective</td>
</tr>
<tr>
<td>• Retrospective</td>
</tr>
<tr>
<td>• Longitudinal</td>
</tr>
<tr>
<td>• Cross-Sectional</td>
</tr>
<tr>
<td><strong>Case-Control</strong></td>
</tr>
<tr>
<td><strong>Case Series</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent lost to follow-up:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Not restricted but should be noted</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample size:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• No restrictions</td>
</tr>
</tbody>
</table>
The systematic search of the literature did not identify any meta-analyses examining the relationship of self-reported pain, self-reported function or walking performance to failure. In the absence of this highest level of evidence, conclusions must be drawn from the conglomerate of cohort studies examining the relationship of interest. A summary of the essential features and findings of the 15 previous studies, found in the literature search as being relevant to the present work, is given in Table III.

Among the 15 studies identified as relevant, none were designed specifically to address the primary question of interest in this thesis, and 8 did not statistically evaluate the association of self-reported pain, self-reported function or walking performance to measures of THR failure (Callaghan, Templeton, Liu, Warth, & Chung, 2006; Giannikas, Din, Sadiq, & Dunningham, 2002; Keener et al., 2003; Y. H. Kim & Kim, 1993; Laupacis et al., 2002; Palm, Olofsson, Astrom, & Ivarsson, 2007; Strom, Kolstad, Mallmin, Sahlstedt, & Milbrink, 2006; Sochart, 1999). Seven studies evaluated the statistical association between pain, function or performance (variably defined) and failure (variably defined) (Bragdon et al., 2007; Gotze et al., 2006; Howie et al., 2007; Kim, Y., 2005; Lautiainen, Joukainen, & Makela, 1994; Soderman et al., 2002; Vervest et al., 2005).

A cross-sectional cohort study of 60 patients with consecutive cementless total hip replacements performed at a single university hospital between 1983 and 1987 by Lautiainen et al. assessed the Pearson correlation between a Modified Mayo pain score and progressive radiolucencies (osteolysis) at a mean time of 5 years after THR (Lautiainen et al., 1994). The mean age was 57.9 years at surgery, 65% were female, and the majority had THR surgery due to OA (76.8%). Investigators concluded that there was no correlation between pain and progressive radiolucencies. Limitations included the evaluation of osteolysis (not validated or used elsewhere), use of a modified, global, clinician-based pain score (discussed below in section 8.1) which, as a group, have been shown to differ significantly from patient centered measures, and a small sample size, limiting statistical power. Large floor effects were noted in the data, making the use of a Pearson correlation coefficient inappropriate due to its underlying assumption of normality.
Kim retrospectively examined a cohort of 100 consecutive THR surgeries performed at the Korean Academic Center at a mean follow-up time of 19.4 years after THR surgery (Kim, Y., 2005). Patients were young, with a mean age of 48.4 years at surgery. The sample was incompletely described so that the sex of participants and the primary diagnosis was unclear. Pain was evaluated using the clinician-based Harris Hip Score and the Visual Analogue Scale to evaluate thigh pain. No association was found between mean linear polyethylene wear and Harris Hip score, but limitations in the study included the method of assessing wear, which was a method developed in-house that utilized AutoCad software where the assumptions used in the calculations were not described (Y. Kim, Kim, & Cho, 2001).

Similar results were seen in another retrospective cohort study done by Gotze et al., at a mean follow-up time of 12.8 years, with 137 cementless THAs from a consecutive series of 231 operations performed between 1986 and 1990 at a single German institution (Gotze et al., 2006). No significant correlations (type not specified) were reported between total Harris Hip score and failure (revision of one or both components of prosthesis) or periprosthetic boney change (Gruen Zones). The primary diagnosis for the cohort was unclear, but patients showed a wide range of ages (24-73) indicating some heterogeneity. Ceiling effects in the Harris Hip score were observed and subscales were not evaluated.

Vervest et al. did a retrospective cohort study that examined 201 patients who had received a unilateral THR due to idiopathic OA in 5 Dutch hospitals (Vervest et al., 2005). Patients had a mean age of 61 at surgery and follow-up was at an average interval of 9 years. Failure was considered to be revision due to aseptic loosening, polyethylene wear determined by manual measurement on x-rays corrected for magnification using known femoral head diameter (Dorr & Wan, 1996), or periprosthetic boney change as defined by the methods of DeLee & Charnley and Johnston. Pain was evaluated using the Charnley and the Harris Hip scores, as well as the Visual Analogue Scale, but it was unclear as to whether this latter measure was hip specific. A Pearson correlation coefficient and Student’s t-test were used to assess associations. Limitations, in addition to the measures used, include low reproducibility of the wear measurements and floor effects with pain data that could obscure any significant associations.

A small sample, single-center cohort study of 28 patients, with 35 THRs (40% female) and with a mean age of 60 at surgery, was performed retrospectively by Howie et al. (Howie et
al., 2007). Pain and function were evaluated using the Harris Hip score, while failure was assessed using computed tomography methods which provided the total volume and annual volume of osteolytic lesions. Polyethylene wear was evaluated using PolyWare computer-based methods. Pain and function were not observed to be associated with the volume of osteolytic lesions. The main limitation of this study was the small sample size made necessary by the precise but expensive computed tomography technique, which limited statistical power.

In a study by Soderman in 2001, a randomly selected subset of individuals from the Swedish National Hip Registry in 9 regional and rural hospitals was evaluated clinically by two orthopaedic specialists (surgeon or a physiotherapist), and radiographically by a third independent observer, 2 to 10 years after THR surgery. Of the original registry cohort, 54% were women and 86% had their replacements due to OA, but the composition of the subset of patients that underwent radiographic and clinical evaluation (344 of 1113 considered patients) was not indicated. Of these 344 individuals, 20% had failed THRs defined as revision surgery for any reason. Pain and function were evaluated on the basis of Harris Hip and WOMAC scores. No statistically significant associations were found, but a borderline unadjusted significant difference in Harris Hip Pain scores between the failure and non-failure groups was observed as evaluated by the Mann-Whitney U test (P = 0.05) (Soderman et al., 2001). Limitations included insufficient power to examine clinical failure in relation to radiographic failure using regression methods, since failure requiring revision was rare (12 of 132). The study did not evaluate polyethylene wear.

In a 2007 retrospective cohort study, Bragdon et al. assessed the association between self-reported pain, self-reported function and radiographic markers of potential THR failure at a mean interval of 6.9 years (range 6-8) after THR in a series of hip replacements with highly cross-linked polyethylene liners which have much lower wear rates than conventional polyethylene liners (Bragdon et al., 2007). Consecutive operations were done at a single institution in 182 patients with 200 THRs. 50% of the individuals in the sample were female but there was no indication of pre-operative diagnosis. Strengths of the study include the assessment of polyethylene wear using steady-state Martell penetration based on multiple time points to get a stable wear estimate. Limitations in the study include the assessment of pain using the WOMAC total score and the Harris Hip score which are subject to confounding or clinician bias (Harris Hip) as discussed below in sections 8.1, 8.4 and 8.5.
8.1 Self-Reported Pain

Of the 7 studies that did test for a statistically significant association of pain with failure, 6 out of 7 concluded that no significant association was evident (Bragdon et al., 2007; Gotze et al., 2006; Howie et al., 2007; Kim, Y., 2005; Lautiainen et al., 1994; Vervest et al., 2005), while the 7th reported an unadjusted borderline significant difference in the Harris Hip score pain subscale between the failure and non-failure groups (Soderman et al., 2001). However, in all 7 studies there were significant limitations in terms of sample size, statistical methodology, or the properties and applicability of the measurement tools. These limitations cast doubt on the significance of the conclusions presented and suggest a need for further work. It was evident, from the review, that no studies evaluated multiple, self-reported measures of pain comprehensively in a single population. Sample sizes were small and there was significant variation within samples, which could explain why associations were difficult to identify, assuming that there are significant relationships between self-reported pain and failure. In certain cases, it was difficult to consider the population as representative, since the diagnosis for THR was not apparent and patients with systemic inflammatory arthritis have a significantly different prognosis compared to the majority of hip replacement patients who have THRs due to OA. There is a propensity to use clinician-based measures for pain or global outcome, meaning measures that clinicians fill out either in presence of a patient or after a clinical visit. In comparison to self-reported measures filled out by patients, these generally differ from the aforementioned clinician evaluations (Musculoskeletal Outcomes Measures and Instruments 2005). In the context of screening tool development in the primary care or at home setting, the author would argue that it is the patient’s report that is of greater interest.

Smaller sample sizes (defined for our purposes as less than 100 participants) were common in the 7 studies (Howie et al., 2007; Kim, 2005; Lautiainen, Joukainen, & Makela, 1994), making study conclusions susceptible to sampling variations and the impact of outliers. It also means that those studies may have lacked the statistical power needed to detect significant changes. An example is the study by Lautiainen et al. in 1994, which was based on 60 patients and where it was concluded that there was no association of radiographic loosening to pain as assessed by a Modified Mayo score. While it is acknowledged that the term ‘small’ is relative, previous limitations with logistic regression using samples of fewer than 100 participants (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996), and limitations in assessing the
relationship of patient-centered factors using regression with samples much larger than 100 (Soderman et al., 2001), support the designation of less than 100 participants as a small sample.

Additional factors to be considered when assessing the suitability of any specific population would include the variable type of implant and the follow-up time after surgery, since implant characteristics may also influence an association between pain and failure. For example, it may be necessary to track newer implants, with highly cross-linked polyethylene liners which have lower wear rates, over longer follow-up periods in order to establish an association with pain because pain may be due to an inflammatory response caused by significant amounts of polyethylene wear (Schmalzried & Callaghan, 1999). This is supported by a 2007 study by Bragdon et al. where the authors looked at patients, who had received a THR with highly cross-linked ultra high molecular weight polyethylene (XLP) liners, at a minimum follow-up time of 6 years and reported no significant correlation between pain, evaluated using the Harris Hip score or WOMAC pain subscale, to polyethylene wear in the context of very limited net penetration.

Statistical limitations in the reviewed studies centered on the frequent use of more than one hip per patient (Bragdon et al., 2007; Howie et al., 2007; Kim, Y., 2005). While this practice is common in the orthopaedic literature, the two hips are not necessarily statistically separate entities. Particularly when related to pain, both hips in an individual would be subject to the same effect-modifying factors central to any one patient (e.g.: perceptions of pain catastrophizing, co-morbidity). The end result could be an over-representation and/or misclassification of pain in the case of logistic analyses which could compromise reported conclusions.

The most significant limitation in this body of prior work is the variety of pain measures employed and the difference in their measurement properties. The variety is a challenge as each measure has individual developmental goals and the concept of pain differs from measure to measure. It is difficult to compare the results from different studies. Moreover, there are significantly different levels of methodological validation for some measures compared to others (e.g.: Merle D’Aubigne-Postel versus WOMAC). This makes it difficult to assess the validity of reported conclusions within the chosen population (Musculoskeletal Outcomes Measures and Instruments 2005)
Table III: Results of systematic literature review.

<table>
<thead>
<tr>
<th>First author Year</th>
<th>Design (Timeline)</th>
<th>Participants</th>
<th>Indicators of THR Failure</th>
<th>Pain, Function or Performance (Measures)</th>
<th>Statistical methods</th>
<th>Test of statistical association of pain, function, or performance with failure (Results)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Bragdon et al., 2007)</td>
<td>Cohort (Retro⁵)</td>
<td>182 (200)</td>
<td>●Polyethylene wear (Martell method)</td>
<td>●Pain</td>
<td>●Multivariate linear regression</td>
<td>Yes (No)</td>
<td>●Medium-term follow-up of highly cross-linked polyethylene in THR implants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50%</td>
<td></td>
<td>●Function</td>
<td></td>
<td></td>
<td>●Age range 22 to 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>---</td>
<td></td>
<td>●Activity</td>
<td></td>
<td></td>
<td>●No correlation between total wear rate and WOMAC, Harris Hip, UCLA activity score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60.1</td>
<td></td>
<td>(WOMAC total and subscales, Harris Hip, UCLA activity score)</td>
<td></td>
<td></td>
<td>●Good distribution of scores</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●Very low penetration rates</td>
</tr>
<tr>
<td>(Howie et al., 2007)</td>
<td>Cohort (Retro⁵)</td>
<td>28 (35)</td>
<td>●Total volume and annual volume of osteolytic lesions determined by CT⁷ scans</td>
<td>●Pain</td>
<td>●Spearman correlation</td>
<td>Yes (No)</td>
<td>●Volume of polyethylene wear is correlated with volume of osteolytic lesions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40%</td>
<td></td>
<td>●Function</td>
<td>Linear regression</td>
<td></td>
<td>●Pain/function are not associated with volume of osteolytic lesions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>---</td>
<td></td>
<td>●Polyethylene wear (PolyWare-computer based)</td>
<td>Mann-Whitney U test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td></td>
<td>●Activity (patient reported 5 point scale)</td>
<td>Kruskal-Wallis test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Palm et al., 2007)</td>
<td>RCT⁴ (Pro⁶)</td>
<td>60 (53)</td>
<td>●Cup migration</td>
<td>●Pain</td>
<td>●Not applicable</td>
<td>No (No)</td>
<td>●Study of wear between cemented low-profile cups and standard cups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>---</td>
<td>●Polyethylene wear (total, 3-dimensional vector, femoral head penetration using stereographic analysis)</td>
<td>●Function - at rest and during activity at 3, 6 and 12 months post-surgery (WOMAC, VAS Pain)</td>
<td></td>
<td></td>
<td>●No differences observed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>---</td>
<td></td>
<td>(WOMAC, VAS Pain)</td>
<td></td>
<td></td>
<td>●VAS higher during unspecified activity</td>
</tr>
<tr>
<td>(Callaghan et al., 2006)</td>
<td>Case control</td>
<td>86 (100)</td>
<td>●Radiographic data (femoral fixation, subsidence, radiolucenties, and stress shielding)</td>
<td>●Clinical (need for revision / re-operation, thigh pain)</td>
<td>●Not Applicable</td>
<td>No (No)</td>
<td>●Medium-term study of extensively coated THR stems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●Age range 18 to 72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●WOMAC scores (only available for one of two groups): mean 13.6 (0-41).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ MACTAR: McMaster Toronto Arthritis Patient Preference Disability Questionnaire
² HRQ of L: Health-Related Quality of Life
³ CS: Cross-Sectional
⁴ RCT: Randomized Controlled Clinical Trial
⁵ Pro: Prospective
⁶ Retro: Retrospective
⁷ CT: Computed Tomography
⁸ ρ: Correlation Coefficient
Table III (Cont’d): Results of systematic literature review.

<table>
<thead>
<tr>
<th>First author Year</th>
<th>Design (Timeline)</th>
<th>Participants N (THR)</th>
<th>Follow-up time in years Mean (Range)</th>
<th>Indicators of THR Failure</th>
<th>Pain, Function or Performance (Measures)</th>
<th>Statistical methods</th>
<th>Test of statistical association of pain, function, or performance with failure (Results)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Gotze et al., 2006)</td>
<td>Cohort (Retro)</td>
<td>117 (137)</td>
<td>12.8 (10-14)</td>
<td>●Radiographic PPBC</td>
<td>●Pain</td>
<td>●Correlation (not specified Pearson or Spearman)</td>
<td>Yes (No)</td>
<td>●Long-term results for the uncemented Lubeck THR in patients aged 24 to 73 ●Large ceiling effect in Harris Hip scores.</td>
</tr>
<tr>
<td>(Strom et al., 2006)</td>
<td>RCT</td>
<td>45 (45)</td>
<td>2 (2-2)</td>
<td>●Radiographic data</td>
<td>●Pain</td>
<td>●Not applicable</td>
<td>No (No)</td>
<td>●Comparison of cemented and uncemented THR in patients aged 41 to 65 ●Looking at stem migration as an indicator of failure ●Males require revision more often and earlier than females</td>
</tr>
<tr>
<td>(Strom et al., 2006)</td>
<td>Cohort (Retro)</td>
<td>82 (83)</td>
<td>8 (7-12)</td>
<td>●Radiographic data</td>
<td>●Pain, Function not assessed</td>
<td>●Not Applicable</td>
<td>No (No)</td>
<td>●Alternate study design of comparison of cemented and uncemented THR as listed above ●Age range 19 to 65 ●No pain or function assessed</td>
</tr>
</tbody>
</table>

1 MACTAR: McMaster Toronto Arthritis Patient Preference Disability Questionnaire  
2 HRQ of L: Health-Related Quality of Life  
3 CS: Cross-Sectional  
4 RCT: Randomized Controlled Clinical Trial  
5 Pro: Prospective  
6 Retro: Retrospective  
7 CT: Computed Tomography  
8 ρ: Correlation Coefficient
Table III (Cont’d): Results of systematic literature review.

<table>
<thead>
<tr>
<th>First author Year</th>
<th>Design (Timeline)</th>
<th>Participants N (THR) % Female % OA Mean age</th>
<th>Follow-up time in years Mean (Range)</th>
<th>Indicators of THR Failure</th>
<th>Pain, Function or Performance (Measures)</th>
<th>Statistical methods</th>
<th>Test of statistical association of pain, function, or performance with failure (Results)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Kim, Y., 2005)</td>
<td>Cohort (Retro)</td>
<td>100 (110) ... ... ...</td>
<td>19.4 (18-20)</td>
<td>Radiographic data</td>
<td>Pain, Function (Harris Hip, thigh pain VAS)</td>
<td>Student’s t-test</td>
<td>Yes (No)</td>
<td>Long-term results for the uncemented porous-coated anatomic THR ● Only total hip scores used ● Wear technique not validated ● Student’s t-test as measure of association not valid</td>
</tr>
<tr>
<td>(Vervest et al., 2005)</td>
<td>Cohort (Retro⁸)</td>
<td>201 (142) 69% 100% 61 (6-12)</td>
<td>Survival (Kaplan Meier – true revision due to aseptic loosening) Polyethylene wear (manual) PPBC (Johnston and DeLee)</td>
<td>Pain (Harris Hip, Charnley, VAS but unclear if hip specific)</td>
<td>Pearson correlation Student’s t-test</td>
<td>Yes (No)</td>
<td>Results after 10 years for the Zweymuller uncemented prosthesis ● Reproducibility of wear measurements was low ● Floor effects with pain data</td>
<td></td>
</tr>
<tr>
<td>(Keener et al., 2003)</td>
<td>Cohort (Pro)</td>
<td>69 (93) 50% 15.9% &lt;50 25 (25-30)</td>
<td>Radiographic imaging and revision surgery in patients with acetabulum aseptic loosening or femoral loosening (no polyethylene wear)</td>
<td>Pain, Function, Performance (WOMAC with all subscales, Harris Hip, 6MWT)</td>
<td>Not applicable</td>
<td>No (No)</td>
<td>Long-term function after Charnley THR ● No significant differences in pain, function or performance between unrevised, unrevised but loose, and revised groups ● Floor effect pain ● Co-morbidity affects function (WOMAC all scales including pain)</td>
<td></td>
</tr>
</tbody>
</table>

¹ MACTAR: McMaster Toronto Arthritis Patient Preference Disability Questionnaire  
² HRQ of L: Health-Related Quality of Life  
³ CS: Cross-Sectional  
⁴ RCT: Randomized Controlled Clinical Trial  
⁵ Pr: Prospective  
⁶ Retro: Retrospective  
⁷ CT: Computed Tomography  
⁸ ρ: Correlation Coefficient
### Table III (Cont’d): Results of systematic literature review.

<table>
<thead>
<tr>
<th>First author Year</th>
<th>Design (Timeline)</th>
<th>Participants (THR)</th>
<th>Follow-up time in years Mean (Range)</th>
<th>Indicators of THR Failure</th>
<th>Pain, Function or Performance (Measures)</th>
<th>Statistical methods</th>
<th>Test of statistical association of pain, function, or performance with failure (Results)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Giannikas et al., 2002)</td>
<td>Cohort (CS³)</td>
<td>66 (71) 45% 91% 55.4</td>
<td>4.8 (2-7)</td>
<td>● Radiographic bone remodeling</td>
<td>● Pain</td>
<td>Not Applicable</td>
<td>No (No)</td>
<td>● Medium-term results for the ABG hip implant in young patients (age 26 to 65)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>● Function (Oxford, Merle d’Aubigne, Harris Hip)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Laupacis et al., 2002)</td>
<td>RCT⁴ (Pro⁵)</td>
<td>250 (250) 48% 100% 64</td>
<td>6.3 (0-9.4)</td>
<td>● Revision surgery of any component for any reason</td>
<td>● Pain</td>
<td>Not applicable</td>
<td>No (No)</td>
<td>● Comparison of cemented and uncemented implants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>● Function (WOMAC, Harris Hip, Merle d’Aubigne-Postel, MACTAR¹)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>● Performance (6MWT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Soderman et al., 2001)</td>
<td>Registry</td>
<td>1113 (1113) 54% 86% 69</td>
<td>--- (2-10)</td>
<td>● Radiographic data (single time point Hodgkinson criteria for loosening in acetabulum or in the stem debonding, stem fracture, cement fracture of 100% circumferential radiolucent line – Methods of Garellick et al 1999 &amp; Mulroy and Harris 1997)</td>
<td>● Pain (subscale analyzed separately)</td>
<td>Mann-Whitney U-test</td>
<td>Yes (Borderline)</td>
<td>● Part II study of THR outcome from data in Swedish Registry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>● Function (Harris Hip, WOMAC)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ MACTAR: McMaster Toronto Arthritis Patient Preference Disability Questionnaire
² HRQ of L: Health-Related Quality of Life
³ CS: Cross-Sectional
⁴ RCT: Randomized Controlled Clinical Trial
⁵ Pro: Prospective
⁶ Retro: Retrospective
⁷ CT: Computed Tomography
⁸ ρ: Correlation Coefficient
Table III (Cont’d): Results of systematic literature review.

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Design (Timeline)</th>
<th>Participants N (THR)</th>
<th>% Female</th>
<th>% OA</th>
<th>Mean age</th>
<th>Follow-up time in years Mean (Range)</th>
<th>Indicators of THR Failure</th>
<th>Pain, Function or Performance (Measures)</th>
<th>Statistical methods</th>
<th>Test of statistical association of pain, function, or performance with failure (Results)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Sochart, 1999)</td>
<td>Cohort (Retro)</td>
<td>163 (235)</td>
<td>19.5 (14-30)</td>
<td>● Total linear wear (manual)</td>
<td>● Pain</td>
<td>● Function (Charnley modification 6 point Merle-D’Aubigne-Postel)</td>
<td>● Not applicable</td>
<td>No (No)</td>
<td>● Relationship of acetabular wear to osteolysis and loosening (age 17 to 39)</td>
<td>● The development of aseptic loosening at an early stage is important</td>
<td>● Radiographic changes are time dependent</td>
<td>● Prediction tool eliminating unnecessary follow-up would be of great benefit</td>
</tr>
<tr>
<td>(Lautiainen et al., 1994)</td>
<td>Cohort (CS(^3))</td>
<td>60 (64) 65% 76.8% 57.9</td>
<td>5 (4-8)</td>
<td>● Progressive radiolucencies (osteolysis) from radiographs</td>
<td>● Pain</td>
<td>● Function (Modified Mayo Score)</td>
<td>● Pearson correlation</td>
<td>Yes (No)</td>
<td>For function, (\rho=0.17) For thigh pain, (p=0.08)</td>
<td>● Clinical and x-ray study of un cemented prostheses</td>
<td>● Compared two types</td>
<td>● Coarse radiographic measures</td>
</tr>
<tr>
<td>(Y. H. Kim &amp; Kim, 1993)</td>
<td>Cohort (Pro)</td>
<td>108 (116) 52% 26.8% 48.4</td>
<td>--- (6-8)</td>
<td>● Radiographic loosening (progressive shift or subsidence) ● Excessive wear (manual techniques) ● Osteolysis ● Heterotopic ossification</td>
<td>● Pain</td>
<td>● Function (Harris Hip)</td>
<td>● Student’s t-test</td>
<td>No (anecdotal)</td>
<td>● Study of un cemented porous-coated anatomic THR</td>
<td>● Compared hip scores between wear and non-wear groups</td>
<td>● No statistical correlation between manual wear and clinical scores</td>
<td>● Anecdotal evidence: thigh pain co-localizes in all cases with loose stems</td>
</tr>
</tbody>
</table>

1 MACTAR: McMaster Toronto Arthritis Patient Preference Disability Questionnaire
2 HRQ of L: Health-Related Quality of Life
3 CS: Cross-Sectional
4 RCT: Randomized Controlled Clinical Trial
5 Pro: Prospective
6 Retro: Retrospective
7 CT: Computed Tomography
8 \(\rho\): Correlation Coefficient
Pain measures used in this area can be broadly classified into clinician-based tools and patient-based tools. Although clinician-based measures were historically thought to have greater validity due to the objective nature of these assessments, recent reports, looking at the disparity of observer and patient-related assessments of pain, and the importance of self-reports when evaluating the construct of pain (Handbook of Pain Assessment, 2001; Jones et al., 2007), imply that clinician-based measures are potentially less relevant for measuring pain than patient-based measures, particularly in the context of pain measures for post-operative surveillance.

Results from surgeon-based measures, for use outside the outpatient health care setting, are inconsistent. For example, in contrast to the group differences which indicated a relationship between pain and outcome in the Finish Hip Registry (Soderman et al., 2001), Kim et al. used the same clinician-based measure (the Harris Hip score) with a younger prospective cohort of 100 patients, at a minimum of 18 years after THR surgery, and showed no association between polyethylene wear and the total score (Kim, Y., 2005). Nevertheless, clinician-based measures appear frequently in the literature. Examples include the Merle D’Aubigne-Postel (1954) Hip score, the Charnley Hip score (1972), and the Harris Hip score (1996) (Bragdon et al., 2007; Howie et al., 2007; Keener et al., 2003; Lautiainen et al., 1994; Sochart, 1999; Vervest et al., 2005).

Another measurement limitation is the relevance of a given measure to the specific sample or population. For example, it has been shown that in the elderly, the VAS measure may be less valid (non-response rates 7-30%) (Gagliese & Melzack, 1997; Herr & Mobily, 1993), and is the least preferred pain intensity measure in the elderly (Herr & Mobily, 1993; Kremer, Atkinson, & Ignelzi, 1981), when compared to other measures of pain intensity such as the Likert format Numeric Rating Scale (Handbook of Pain Assessment, 2001). Nevertheless, Vervest et al. used the VAS pain measure in a sample with a mean age of 61 at surgery (72 at follow-up) (Vervest et al., 2005). While there is no discussion surrounding the impact of this choice on missing data, significant loss of clinical data was observed. The authors concluded that the pain measurements did not correlate with wear (evaluated manually) but if patients were unable to properly complete the VAS measure, this could have significantly influenced those findings. While this may be one reason why the results showed no relationship of pain to radiographic wear, it is important to note the authors’ observation that the reproducibility of the
manual wear measurements was low. Contrary to the claims of Vervest et al. (Vervest et al., 2005), computer techniques have been frequently shown to provide superior precision and reproducibility in comparison to manual techniques (McCalden et al., 2005).

Despite all these challenges, there is limited support for some form of relationship between pain and radiographic failure. The Swedish National Hip Registry (Soderman et al., 2001), demonstrated a significantly higher proportion of patients with pain (Harris Hip pain subscale) who showed indications of radiographic failure (Hodgkinson’s criteria) in either the cup or the stem of the implant, even though floor effects, present in the majority of pain data, would tend to obscure positive results (Smythe & Bogoch, 2008). Failure of the acetabular component was defined as Hodgkinson Type 3 with 100% circumferential radiolucent lines, while failure of the femoral component comprised debonding, stem fracture, cement fracture or a 100% circumferential radiolucent line in the femoral component (Hodgkinson et al., 1988).

8.1.1 Factors Influencing the Pain-Failure Relationship

No studies reviewed, addressed or controlled for pain catastrophizing even though recent literature has underscored the importance of such coping behaviours in the conceptualization of pain. Pain catastrophizing is defined as “an exaggerated negative orientation toward pain stimuli and pain experience” (Handbook of Pain Assessment, 2001; Keefe et al., 2001; M. J. L. Sullivan, Bishop, & Pivik, 1995).

It can include magnifying the consequences of pain, ruminating excessively about pain, and feeling helpless in managing or coping with pain. This exaggerated response to pain has been used to explain individual differences or variations in the levels of pain and distress reported in musculoskeletal disease. Specifically, those with high levels of pain catastrophizing may be more likely to over-report pain intensity or impact. A growing body of literature associating this construct with pain and OA implies that it may significantly influence the measurement of pain in THR patients (Keefe et al., 2000; Somers et al., 2009; Somers, Keefe, Godiwala et al., 2009; Somers, Keefe, Godiwala, & Hoyler, 2009). Pain catastrophizing could act as an effect modifier (Kratz, Davis, & Zautra, 2007) of self-reported pain causing an over or under representation of pain in patients with various levels of structural change. Therefore, to determine the relationship between pain and THR failure independent of perceptual differences in pain, it is important to control for these factors when examining the pain-failure relationship.
A significant result would only be meaningful in the context of a candidate THR surveillance tool if the relationship remained significant when the association was adjusted. While the impact of pain catastrophizing on objective function in THR patients has yet to be explored, pain catastrophizing in total knee replacement patients has been shown to be associated with disability pre and post-surgery, even after controlling for the variance in levels of pain intensity, depression and anxiety (M. Sullivan et al., 2009).

Co-morbidity has been shown to be strongly predictive of THR failure, irrespective of the follow-up period (Johnsen, Sorensen, Pedersen, Lucht, Soballe, & Overgaard, 2006). Co-morbidity is associated with increased pain, decreased function and decreased performance in patients with OA of the hip or knee, independent of age or gender (van Dijk et al., 2008). Co-morbidity has received variable levels of attention in the 15 studies considered for review and should be considered in future evaluations of the pain/failure relationship.

Time after hip replacement surgery should also be considered as a modifying factor in the pain relationship with failure. There is a proportional relationship between time after surgery and the amount of polyethylene wear. Since it has been suggested that there is a time-varying relationship of pain development with the progression of OA, which is expected to be mirrored in the post-surgical population, this has the potential to influence the pain to wear relationship if not correctly addressed.

### 8.2 Self-Reported Function

Self-reported function refers to a patient’s report of any impairment or limitation on their ability to perform various activities of daily living. Of the 15 studies subject to full review, 14 examined the relationship of self-reported function to failure, defined variably as: several definitions of periprosthetic boney change, osteolysis and loosening, polyethylene wear, or revision surgery of any component. None of the studies reported a significant association (Callaghan et al., 2006; Giannikas et al., 2002; Keener et al., 2003; Kim and Kim, 1993; Laupacis et al., 2002; McGrory et al., 1996; Palm et al., 2007; Soderman et al., 2001; Strom et al., 2006). In general, measures employed to evaluate function used the same composite outcome measures discussed above (Western Ontario and McMaster Osteoarthritis Index (WOMAC), the Charnley Hip score, the Harris Hip score 1969, the HOOS and the Oxford Hip score) all of which contain a subscale assessing self-reported function with varying weights and
compositions primarily focused on activities of daily living. Frequently, there were no details provided regarding subscale evaluation so that it remains unclear whether function alone did not relate to failure or if the combined outcome construct did not relate. Moreover, previous discussion on the large influence of pain on function, particularly in the context of the WOMAC measure, suggests that it is not function alone being evaluated, but rather the combined construct, heavily influenced by pain (G.A. Hawker et al., 2008; van den Akker-Scheek et al., 2008), a fact compounded when the subscales are not evaluated separately as is frequently the case.

It is has also been suggested that these measures of function may assess a patient’s experience in performing a task instead of their actual ability to perform it (Stratford & Kennedy, 2006), meaning that two patients might walk the same distance in the same amount of time (equal performance score) but the individual with perceived difficulty would have higher self-reported functional impairment. A growing body of literature highlights the poor correlation of self-reported function with walking performance (van den Akker-Scheek, Zijlstra, Groothoff, Bulstra, & Stevens, 2008). This suggests that measures of both function and performance are required to obtain a full picture of functional ability and may suggest that performance, rather than function, may be more reliably related to THR failure, as it is not confounded by pain or a patient’s subjective experience (Stratford & Kennedy, 2004; Stratford & Kennedy, 2006). Nevertheless, while performance is important, it may also be artificial and of limited relevance to daily life, a view that supports the complementary nature of the two constructs.

8.3 Walking Performance

Of the studies under review, few evaluated objective performance as it relates to THR failure. The most common performance measure used in post-operative THR patients is the six minute walk test (6MWT), as it is reflective of the functionality required in daily life. This measure is considered to be objective as it does not appear to be influenced by self-perception of ability (Kennedy, Stratford, Wessel, Gollish, & Penney, 2005). Only two of the 15 studies under review evaluated this concept using the standardized six minute walk test (Keener et al., 2003; Laupacis et al., 2002). No association was found between performance and failure (variably defined) but the limitations in sample size, population characteristics and statistical methodology, as previously noted, were also evident.
8.4 Assessment of THR Failure

The variety of criteria used in the definition and assessment of THR failure is another possible explanation for the inconsistent relationship of pain, function and performance to THR failure or its markers. This issue is not unique to the pain-failure relationship but is a continuous challenge within orthopaedics and even medicine as a whole. It makes comparisons challenging or even impossible. Therefore, it is necessary to perform a systematic evaluation, within a single population, to reliably evaluate the relationships of pain, function and performance to relevant markers of potential THR failure.

The impact of variable definitions of failure can be seen when examining the findings of Lautianinen’s study in 1994 (Lautiainen et al., 1994) compared to the findings of other studies, such as those of Sochart and Bragdon. Essentially, the three studies cannot be compared as the definition of failure used in the Lautiainen study is very stringent and will therefore underestimate a relationship relative to other definitions, while Bragdon and Sochart used a more inclusive definition, which tended to pick up more of the borderline cases. Radiographic failure in the Lautiainen study was identified by loss of points determined by measurements of osteolytic lesions greater than 2 mm or component migration (both later signs of failure in a cohort 5 years post-surgery). As this level of osteolysis is quite large with improved low bearing surface wear, and may rarely be present 5 years post-surgery, the tendency of this measure would be to underestimate boney change at this time period, which is already underestimated by the use of radiographic technology.

By comparison, others studies (Bragdon et al., 2007; Howie et al., 2007; Keener et al., 2003; Lautiainen et al., 1994; Sochart, 1999; Vervest et al., 2005) considered all aspects of bone change (osteolysis, osteopenia, cortical hypertrophy, heterotopic ossification and pedestal). These differences in evaluation criteria make it difficult to draw firm conclusions based on prior work (Y. H. Kim & Kim, 1993; McGrory et al., 1996; Sochart, 1999; Soderman & Malchau, 2001; Soderman et al., 2001).

8.5 Newer Measures of Pain

Newer measures of pain are needed to evaluate post-operative THR patients. In the development of screening tools for use in an at-home or primary care setting, surgeon-based
measures are of only limited use. Existing patient-centered pain measures such as the WOMAC pain subscale may be confounded by function and do not necessarily reflect the pain experienced by patients (G.A. Hawker et al., 2008; van den Akker-Scheek et al., 2008).

In 2007, the measure for Intermittent and Constant Osteoarthritis Pain (ICOAP) was developed by Hawker et al. using qualitative methods to capture the pain experience in hip and knee OA patients (G.A. Hawker et al., 2008). The concept of the measure was to examine OA pain taking into account the impact of pain on mood, sleep and the activities of daily life. It was developed for individuals with symptomatic hip and knee OA where preliminary data revealed a different, time-varying pain profile moving from intermittent early pain when compared to more constant pain with disease progression (G.A. Hawker et al., 2008). It is expected that this trend will hold in a post-operative population (A. Davis et al., unpublished data). Factor analysis has validated two separate sub-scales (G.A. Hawker, French, Davis, & Dieppe, 2008) corresponding to intermittent and constant pain. Focus groups endorse a temporal evolution of pain progression where intermittent pain is more frequently present early in disease progression and, therefore, is more likely to be seen concomitant with early indicators of THR failure (G.A. Hawker et al., 2008). The use of this measure may provide a more sensitive measure of pain in a post-operative THR cohort.

8.6 Overview of the Association of Self-Reported Pain and Function, and Walking Performance with THR Failure

None of the 15 evaluated manuscripts investigated the association of pain, function and performance with THR failure as their primary objective. What little literature exists on the relationship of these indicators to failure in THRs does not constitute a comprehensive evaluation of this topic as several key limitations are apparent: 1) No studies looked concurrently at a wide variety of measures of pain and function in the same sample, 2) Most samples were restricted to a single or a few surgeons with a specific type of implant which may not be representative of the entire THR population, 3) The majority of studies were based on small sample sizes (<100), sometimes including rheumatoid arthritis and other related conditions with different pathogeneses and outcomes, 4) There is considerable heterogeneity in terms of predictors, exposures and outcomes employed, in addition to significant variation in the follow-up periods (i.e.: cross-sectional pain and functional deficits at failure) making reliable comparisons challenging, 5) Frequently, the statistical analyses included two hips per patient but
did not account for the lack of independence of these data points, and 6) Since 2007, new tools for measuring pain and function in patients with OA have been developed and the technology for measuring wear has advanced, potentially allowing discovery of underlying associations that might previously have remained undetected. Therefore, re-examination, using improved measures, of the relationship of pain, function and performance to radiographic indicators of failure is required.

9 Summary

Surveillance for THR patients is highly recommended as undetected failure is associated with significant morbidity and even mortality (Paprosky et al., 1995; Williams et al., 2004). Current methods of surveillance, which require regular follow-up with an orthopaedic surgeon or other orthopaedic specialist, are unsustainable, with the ever increasing number of new THRs being performed without the proportional addition of trained personnel. Therefore, alternative measures of surveillance that can be performed at home or in the primary care setting, such as self-report measures, are required. Self-reported pain and function and performance measures of function can be reliably measured, but have yet to be consistently associated with THR failure or early radiographic indicators. One potential explanation for these inconsistencies might be the insensitivity of commonly used measures of pain, function, or performance and/or previous technological limitations in evaluating early radiographic indicators of failure such as polyethylene wear and periprosthetic boney change. As newer, and potentially more sensitive, measures of pain are now available, methods to evaluate wear measurement have advanced, and additional relevant covariates have been identified, the present study was initiated to re-examine the hypothesis that pain, function and performance relate to early indicators of radiographic failure, such as wear and periprosthetic boney change.
Chapter 3 Objectives & Hypotheses

Section 10 outlines the overall objectives and specific primary and secondary objectives of this study. Section 11 outlines the detailed hypotheses.

10 Objectives

10.1 Overall Objective

To comprehensively assess the cross-sectional associations of self-reported pain, self-reported function and walking performance with two early radiographic markers of potential THR failure (polyethylene wear and periprosthetic boney change) in a single sample as a first step towards development of a potential tool for post-operative surveillance for THR failure in patients who have had a hip replacement for OA.

10.2 Specific Primary Objectives

1. To evaluate the crude and adjusted (for covariates/confounders) association of self-reported measures of pain to polyethylene wear, an early radiographic marker of potential THR failure.

2. To evaluate the crude and adjusted association of self-reported measures of physical function to polyethylene wear.

3. To evaluate the crude and adjusted association of an objective measure of physical function, the six minute walk test, to polyethylene wear.

4. To determine which patient-centered construct: self-reported pain, self-reported function or walking performance, or combination thereof, best discriminates patients ‘at-risk’ for THR failure, as defined by those with a high annual polyethylene wear rate (≥ 0.1 mm/year versus < 0.1 mm/year).

10.3 Specific Secondary Objectives

5. To evaluate the crude and adjusted (for covariates/confounders) association of self-reported measures of pain to periprosthetic boney change, an early radiographic marker of potential THR failure.
6. To evaluate the crude and adjusted association of self-reported measures of physical function to periprosthetic boney change.

7. To evaluate the crude and adjusted association of an objective measure of physical function, the six minute walk test, to periprosthetic boney change.

8. To determine which patient-centered construct: self-reported pain, self-reported function or walking performance, or combination thereof, best discriminates patients ‘at-risk’ for THR failure, as defined by those with at least some periprosthetic boney change (variably defined: see methods).

9. To assess the level of agreement between two radiographic markers of potential THR failure, polyethylene wear and periprosthetic boney change, when classifying individuals ‘at-risk’ for THR failure.
11 Hypotheses

11.1 Specific Primary and Secondary Hypotheses

1. $H_01$: Pain, Function and Performance will be significantly associated with radiographic markers of potential THR failure (annual polyethylene wear rate of $\geq 0.1$ mm/year or at least some periprosthetic boney change) $(0.5<\rho<1$, odds ratio (OR) $>1)$, even after adjustment for other factors (time post surgery, pain catastrophizing, etc.)

2. $H_02$: Self-reported pain in general and specifically, the Intermittent ICOAP pain score will have the strongest association with radiographic markers of potential THR failure (annual polyethylene wear rate of $\geq 0.1$ mm/year or at least some periprosthetic boney change) as indicated by the largest value of association (largest $\rho$/OR: $0.5<\rho<1$, OR $>1$).

3. $H_03$: Walking Performance will be more strongly associated with radiographic markers of potential THR failure (annual polyethylene wear rate of $\geq 0.1$ mm/year or at least some periprosthetic boney change), than measures of self-reported function indicated by larger measures of association (larger $|\rho|$/OR).

4. $H_04$: Polyethylene wear and periprosthetic boney change will show fair to moderate agreement on the classification of ‘at-risk’ individuals for THR failure, as determined by a kappa coefficient adjusted for chance between 0.2 - 0.6.
Chapter 4 Methods

12 Methods

12.1 Study sample

12.1.1 Sampling Frame

This cross-sectional study was nested within an ongoing prospective cohort study that is examining physical activity following total hip replacement surgery. Therefore, potential participants for this thesis project were recruited from among participants in the larger study. Eligibility criteria for the larger study included English speaking adults 80 or less at the time of surgery, with primary or secondary OA. Patients with metastatic cancer, Paget's disease, Lupus, trauma or previous fracture in the study hip, limited cognitive ability (dementia), or those with significantly impaired physical function resulting from inflammatory arthritis or systemic disease requiring disease modifying medications were excluded.

Four hundred and eighty-eight individuals who had received a primary THR due to OA between January 2001 and June 2004 at Sunnybrook Health Sciences Center (SB) and/or Holland Orthopaedic and Arthritic Institute (HC) were identified from an eligibility list provided by Health Data Records and contacted by research staff, in ascending order of surgery date, in order to obtain verbal consent for participation in a research study. Participants were informed of the goals of the prospective study, which were to develop physical activity recommendations for patients with THR. If potential participants provided consent, they were screened by telephone for eligibility in the current study and booked for a research appointment and clinical follow-up with their surgeon or his designate, an advanced practice physiotherapist specialized in orthopaedics. At the research appointment, details of image history and implant characteristics of potential participants were obtained from medical records after the participants signed a written informed consent document (Appendix 1). From among those eligible for this larger study, we selected, for participation in the thesis study, those who had appropriate implant and radiographic criteria as detailed below.
12.1.2 Inclusion Criteria

Patients who had received a THR due to a primary diagnosis of osteoarthritis (OA) or a diagnosis of OA secondary to congenital hip dysplasia and/or avascular necrosis were included, provided that they were less than 80 years of age at the time of their THR and could read and comprehend English so that they could fill out the questionnaires and provide informed consent. Potential respondents were required to have a standard THR implant (metal backed acetabular cup, polyethylene (conventional or highly cross-linked) liner, and a metal or radio-opaque ceramic femoral head) in the study hip.

Only patients with two complete sets of pelvic x-rays, the first occurring 28-90 days after surgery and the second within ±90 days of the research appointment, were included. Complete sets of images were defined as a bilateral anterior posterior (AP2) pelvic view, a unilateral anterior posterior view of the replaced hip (AP1) and a lateral image of the replace hip (LAT).

12.1.3 Exclusion Criteria

Patients with conditions known to affect the survival of the THR and/or restrict levels of physical activity (metastatic cancer, Paget's disease, trauma or previous fracture of the proximal femur or acetabulum in the study hip) were excluded. To address the issue of statistical independence of data points, only one replaced hip was included per study respondent in instances where a potential respondent had had two hips replaced within the study time-frame.

To identify the study hip in individuals with two eligible prostheses, participants were asked to identify the hip that was the ‘most painful or troublesome’. In the event that patients could not identify one hip over the other, the patient, in combination with medical records, identified the hip with the oldest prosthesis. If a patient had had both hips replaced during the same surgical procedure, research personnel took the right hip by default. The rationale for this choice was that investigators wanted to identify hips at risk for failure. Since joint use is one potential predictor of failure, and the majority of the population is right leg dominant, it was assumed that, if there was a difference, the bias would be toward those hips with marginally more use and hence risk for failure. Patients with alternative bearing surface implants (metal-on-metal, ceramic on ceramic, or radio-translucent ceramic on polyethylene) or an all polyethylene acetabular cup cemented directly to bone, were excluded from assessment because the Martell
method, used to evaluate the primary outcome (radiographic wear), cannot reliably evaluate wear in these implants.

12.2 Assessments

Individuals who met the eligibility criteria, and who agreed to participate, attended a combined research and clinic appointment, no longer than two hours in duration, where x-rays were taken, the participant met with their surgeon or his designate at their location of practice (HC or SB), and with the research staff. The majority of participants were seen by a single member of the research team (JMC) with four other members available for support in the event of scheduling conflicts. The x-ray and clinic appointments followed standard practice guidelines for post-operative surveillance of THR patients. During the research appointment, demographic information was collected, height and weight measurements were made, responses to interviewer-administered questionnaires were recorded and participants completed a standardized six-minute walk test. Prior to the appointment, research staff met with participants so that they could read, ask questions, and sign the informed consent document approved by the Sunnybrook Research Ethics Board. Participants volunteered their time freely and no honorarium was provided.

12.3 Ethics

This study was approved by the Research Ethics Board at Sunnybrook Health Sciences Center. Amendments to permit the collection of additional radiographic data were evaluated and approved in February and June 2008.

12.4 Demographics & Surgical Characteristics

Demographic information (Appendix 2), including age at THR surgery, sex, marital status and employment status, was collected at the research appointment. Researchers verified Health Data Records information regarding the surgical characteristics of each participant, including primary diagnosis, surgery date and hip replaced (right or left), while collecting information (Appendix 3) about the design, the bearing surface, the materials, and serial numbers of the patient’s implant and its components from the patient’s medical records.
12.5 Assessment of Self-Reported Pain, Function, & Walking Performance

Pain and functional difficulties were assessed by means of the following standardized interviewer-administered and/or self-administered questionnaires. Walking performance (objective function) was assessed using a standardized performance measure as described below.

12.5.1 The Western Ontario and McMaster Osteoarthritis Index (WOMAC).

WOMAC 3.1 (Appendix 4) is a 24-item health status measure with 3 domains. It is the most widely used OA outcome measure for the hip and knee and is generally considered to be reliable, valid and responsive to change in the condition of hip OA patients in both pre and post-surgical populations (Soderman & Malchau, 2000; Soderman & Malchau, 2001; Soderman et al., 2001; Stucki et al., 1998; Thumboo, Chew, & Soh, 2001; J. G. Wright & Young, 1997). The WOMAC is composed of: a pain subscale with five items (WOMAC-P), a stiffness subscale with two items (WOMAC-S), and a physical function subscale with 17 items (WOMAC-PF). An aggregate score can evaluate global participant status (WOMAC-G). All items are evaluated on a 0-4 Likert scale with response options: 0 = none; 1 = mild; 2 = moderate; 3 = severe; and 4 = extreme. Higher scores correspond to higher levels of symptoms or disability. Questions focused on the participant’s sensations of pain, restrictions and ability to perform certain activities of daily living such as walking, stair climbing, putting-on and taking-off socks or stockings, and doing both heavy and light domestic duties. The questionnaire was either self-administered or interviewer-administered by research personnel depending on the availability of appropriate personnel. In order to match the timeframe of other measures of pain and function employed in this study, participants were instructed to complete the questionnaire by recalling their experience during the previous week (Bellamy, 1989).

Each subscale was considered, operationally, as an ordinal variable but was also categorized for the purposes of logistic regression into clinically relevant categories. For the pain subscale WOMAC-P, participants with scores ≤ 7 reported, on average, less than moderate pain on each of the five items in the subscale and were considered to have a low level of pain compared to their higher-pain counterparts. This score corresponds to an average between ‘mild’ and ‘moderate’ pain on each of the five items or implies only mild pain on all but one or two items on the subscale. For the physical function subscale, WOMAC-PF, participants with scores
≤ 17 were considered to have no discernable difficulty compared to their higher-scoring counterparts. This score is equivalent to an average of ‘mild difficulty’ or less on each of the 17 items in the subscale.

12.5.2 The Numeric Rating Scale (NRS) for Arthritis Pain

The Numeric Rating Scale (Appendix 4) is an 11-point pain intensity rating scale. Participants were instructed to ‘rate their arthritis pain on a scale from 0 to 10, 0 being ‘no pain at all’ and 10 being the ‘worst pain possible’. No time frame or specification with respect to the study hip was provided due to constraints of the sampling frame (requirement for a larger cohort study). This scale has been used extensively in pain research, and in the THR population, and has been shown to perform well in the elderly (Handbook of Pain Assessment, 2001).

12.5.3 The Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP)

This questionnaire (Appendix 4) contains 11 questions evaluating the pain experience in OA. It was developed from qualitative data and comprises two theoretical domains of pain that are clinically relevant: constant pain (5 items) (e.g.: “In the past week, how intense has your constant hip pain been?” or “In the past week, how frustrated or annoyed have you been by your constant hip pain?”) and intermittent pain (6 items) (e.g.: “In the past week, how intense has your most severe hip pain that comes and goes been?” or “In the past week, how frequently has this hip pain that comes and goes occurred?”). The questionnaire was interviewer-administered in person or by phone. Participants responded on the basis of the pain experienced in their study hip within the last week. The ICOAP has been shown to be both reliable and valid in a pre-surgical OA population suffering from moderate to severe OA (G.A. Hawker et al., 2008; G.A. Hawker et al., 2008; G.A. Hawker, French et al., 2008). It has also been recently shown to be reliable, valid and responsive in both pre and post-total joint arthroplasty. Scores were normalized to 0-100 for each of the two subscales and the total score. As above, for the purpose of logistic regression, both the normalized subscales and the total score were used, operationally, as dichotomous as well as continuous variables. Participants with an average of ‘none’ to ‘mild’ pain (scores ≤ 25/100) were compared to those with an average of ‘moderate’ to ‘extreme’ pain (scores > 25/100).
12.5.4 The Six Minute Walk Test (6MWT)

The Six Minute Walk Test (Appendix 2) is a validated performance measure in geriatric populations and was developed as a proxy to assess participant capacity to perform the activities of daily living. Participants were asked to walk for six minutes at a self-selected pace covering as much distance as possible. Rest breaks were permitted, as required, without stopping the clock, and the number of breaks was recorded. The test has proven to be effective in the THR population, with no apparent differences by implant type. Results are typically reported as distance walked in meters; thus, higher values indicate better physical functioning (R. Barrack et al., 2006; Kennedy, Stratford, Wessel, Gollish, & Penney, 2005).

It has been recently suggested, in the cardiovascular and lung literature, that outcomes for the 6MWT may be more useful measures of impairment if compared to the predicted distance a comparable, but healthy, individual would walk under the same conditions. (Balashov et al., 2008) Therefore, in addition to the distance walked in meters, a normalized score was calculated and analyzed as an indicator of walking performance. Calculation of the normalized score involved taking the distance walked in meters relative to the predicted distance walked in meters by a healthy control matched in age, sex, height and weight. Of the available predictive equations, Enright’s equation, derived in a healthy elderly population in the United States, was chosen to calculate the denominator due to the similarity of the population and the walk test protocol used by Enright (Enright & Sherrill, 1998; Kennedy, Stratford, Wessel, Gollish, & Penney, 2005; Troosters, Gosselink, & Decramer, 1999) to that of the current study.

12.5.5 Six Minute Walk Test Pain

As the NRS is designed to contextually measure pain intensity, a second, more specific pain NRS was administered to participants, immediately upon completion of the six-minute walk test. Participants were asked to “rate their hip pain” referring to their replaced hip “on a scale of 0-10 where 0 = no pain at all and 10 = the worst pain possible” (Appendix 2). Participants with no pain (scores 0/10) were compared to participants with at least some pain (score ≥1) for the purposes of logistic regression.

12.6 Potential Covariates

The two following measures were also self-administered or administered by an interviewer.
12.6.1 Self-reported Co-morbidity Index

In 1996, Katz et al. (Katz, Chang, Sangha, Fossel, & Bates, 1996) demonstrated that self-reported co-morbidity could be reliably used as an efficient alternative to chart abstraction in order to assess co-morbidities for research purposes. This precipitated the development of a number of targeted, self-report co-morbidity indices. The co-morbidity index employed here (Appendix 2) was developed in the mid 1990’s by G.A. Hawker and colleagues, specifically to evaluate co-morbidities that might impact lower extremity symptoms and disability in OA. This measure has been proven to be a reliable and valid method of assessing co-morbidities for presurgical lower extremity OA patients (all ranges of disease) in an established and previously described cohort of hip and knee OA sufferers (G.A. Hawker, Wright, Glazier, Coyte, Harvey, Williams, & Badley, 2002). Participants were categorized, for descriptive purposes, into three groups based on the number of co-morbidities reported: one or less, exactly two, or three or more.

12.6.2 The Pain Catastrophizing Scale

The Pain Catastrophizing Scale (Appendix 4) is a 13-question measure assessing participant perceptions of pain. The three-domains are pain magnification (e.g.: “There is nothing I can do to reduce the intensity of the pain”), rumination (e.g.: “I worry all the time about whether the pain will end”), and helplessness (e.g.: “I feel I can’t go on”). All items were rated on a 0-4 Likert scale, looking at the recalled frequency of experiencing a particular thought or feeling (0 = never, while 4 = all the time). Total individual scores were calculated as the sum of the scores for each question (maximum = 52) to provide a single covariate for adjustment within a multiple logistic regression analyses. Participants were instructed to consider any previous painful experience, and rate their perceptions in that context, but not to focus necessarily on arthritis pain (Sullivan, Bishop, Pivik 1995).

12.7 Assessment of THR Wear

12.7.1 The Martell Method of Wear Analysis

The purpose of the wear analysis was to accurately determine the amount and direction of femoral head penetration into the plastic liner between the 28-90 day post-operative and most recent radiographs. This was accomplished using the Martell method (Martell & Berdia, 1997).
As indicated in Figure 3, the Martell method utilizes a computer-based, semi-automatic edge detection software program called the Hip Analysis Suite (University of Chicago, Chicago, Ill) (Martell & Berdia, 1997). The underlying premise of the Martell method is to allow the computer to determine the best fit circles around the femoral head and acetabular shell and then calculate the direction and rate of wear based on the relative change of these circles between Time Point 1 and Time Point 2 using paired x-ray images. The calculation is semi-automatic as it requires the operator to note relevant landmarks on the image (the most inferior point of the ischial tuberosities, as well as the region of the femoral head and acetabular shell) so that the calculations are oriented to a particular image and normalized using pixilation within the image and outside the region of interest. The method is similar, in theory, to digital callipers but the limits of resolution are much better, at 0.025 mm. Once a region of interest is selected, a two-step image conversion process maximizes contrast in the focal region and normalizes the exposure to the light conditions. Following this transformation, the program uses mathematical techniques to establish the best fit circles. The distance, total linear wear, denoting relative change between the two circles between time points is calculated on the basis of pixel differences and converted to mm based on the known diameter of the femoral head. The direction of wear, the ‘wear vector’, is calculated from the center of the femoral head in the direction of femoral head displacement. Total volumetric wear, the estimated volume of polyethylene released from the bearing surface, is calculated using the wear vector and a standard volumetric formula based loosely on the volume of a sphere.

Total linear penetration (linear wear) and total volumetric wear were calculated between the first post-operative radiograph and the last follow-up radiograph (in mm or mm$^3$ respectively) and normalized to the duration of each patient’s radiographic follow-up to give a measure of annual linear or volumetric wear rates. Annual linear wear was used as the primary outcome because thresholds corresponding to significant osteolysis have been well established for this measure. Based on these thresholds for the purposes of logistic regression, participants were categorized into a high (> 0.1 mm/year) versus low (≤ 0.1 mm/year) annual linear wear rate groups. The less conservative osteolysis threshold for polyethylene wear ≤ 0.1 mm/year was used because there were insufficient participants with the more conservative ≤ 0.2 mm/yr wear rate in the study sample.
In the Martell method, the centre of the cup circle is considered to be a reference point. The position vector for the femoral head is defined as the vector running from the centre of the cup circle to the centre of the head circle and this vector will change with polyethylene wear.

**Figure 3:** Determination of 2-dimensional linear wear using the Martell method. In the enhanced radiograph, a hip analysis software package has been used to fit circles and ellipses to the boundaries of a THR implant comprised of a metal acetabular cup, a polyethylene liner and a metal or radio-opaque ceramic femoral head. As indicated in the sketch, linear wear is determined by the vector change in femoral head position from x-rays taken soon after surgery (Time Point 1) and 4 to 8 years post-surgery (Time Point 2). Distance calibration within and between x-rays is provided by the known dimensions of the femoral head while angles are referred to the horizontal defined by the ischial tuberosities. Information on the inclination and anteverision of the acetabular cup can be obtained from the geometry of the elliptic image of the rim.
12.7.2 Acquisition and Preparation of x-rays

Two 3-image series of pelvic X-rays, including a bilateral anterior posterior (AP2) image, a unilateral image focusing on either the right or left hip (AP1), and a lateral (LAT) image of the study hip were collected, if available. The first time point corresponded to an early post-THR series occurring 28 to 90 days after the THR surgery. The second time point was the series obtained at the research appointment; 4 to 8 years post surgery. The required images, high quality Digital Imaging and Communications in Medicine (DICOM) images are equivalent to high quality JPEG images but also carry additional patient information. The DICOMs were either downloaded from a hospital database and anonymized using a freeware version of IMAGE Information Systems Ltd. DICOM Anonymize software (London, UK), or hard copy films were obtained from the medical imaging department and digitized to DICOM format using the Vidar Film Scanner. These images were anonymized prior to digitization using temporary blackout labels to cover identifying image tags. Two bilateral anterior-posterior x-rays with good image quality were required for inclusion in the Martell analysis. Acceptable image quality was determined by a single trained research technician with experience performing Martell analyses (McCalden et al., 2009). No patients were excluded on the basis of image quality.

12.7.3 Evaluation of Periprosthetic Boney Change

An independent orthopaedic surgeon, who was blinded to the research question and pain scores, evaluated the two anterior posterior (bilateral and unilateral) and lateral radiographs at the two time points. This evaluator looked for the development or progression of radiolucencies or other indications of bone remodelling around the hip prosthesis using the method of DeLee and Charnley (DeLee & Charnley, 1976), as well the method of Gruen et al. (Gruen, McNeice, & Amstutz, 1979) and the extension developed by RC Johnston et al. (Johnston et al., 1990) for use with lateral radiographs. Specifically, the evaluator was looking for the presence or absence of developing and/or progressing osteolysis, osteopenia, cortical hypertrophy, pedestal, or heterotopic ossification within each of the 14 zones shown in Figure 4. These methods (universally applied in various forms across the literature) are considered valid and reliable and are employed frequently in the post-operative THR population (Soderman & Malchau, 2001). However, inter-observer reliability, assessed using an intra-class correlation coefficient to evaluate THR stems (using aseptic loosening as the outcome), demonstrated only moderate inter-
observer reliability (ICC = 0.47) overall, and poor agreement (ICC = -0.10) when considering uncemented stems alone. Despite what could be seen as a limitation when compared with other radiographic methods, plain x-rays show the best overall scores, including high sensitivity of 81.3% (63.6–92.8), acceptable specificity of 73.9 % (58.9–85.7) and a better than chance positive predictive value of 68.4% (54.8–85.8) with a good negative predictive value of 85.0% (70.2–94.3) (Temmerman et al., 2006). Similar results, based on a 2007 systematic review, were seen with acetabular components, again using aseptic loosening as the outcome (Temmerman et al., 2007).

Each participant was categorized into a high versus a low-risk group for THR failure in three different ways: (1) The most inclusive definition defined PPBC as any evidence of osteolysis, osteopenia, cortical hypertrophy, pedestal, or heterotopic ossification in any zone (yes/no), (2) The second definition excluded consideration of osteopenia and PPBC was therefore defined as any evidence of osteolysis, cortical hypertrophy, pedestal, or heterotopic ossification in any zone (yes/no), and (3) As many patients had HO alone, the third definition of PPBC considered all patients with heterotopic ossification of any Brooker’s grade. While ideally, each element of PPBC would be evaluated independently, few individuals with any evidence of PPBC meant that there was insufficient statistical power to evaluate any bone remodelling, apart from heterotopic ossification, independently.
Figure 4: DeLee Zones for evaluating periprosthetic boney change (PPBC) in the acetabulum (I-III). Gruen Zones for evaluating PPBC in the femur (Zones 1-7). Johnston Zones for evaluating PPBC in the lateral view of the femur (Zones 8-14).
12.8 Statistical Analysis

12.9 Descriptive Statistics

Data for each variable were summarized using counts, percentages, measures of central tendency (mean, median) and measures of sample variation (sample standard deviation, range) within the study sample overall, and for the high-wear group and the low-wear group separately. Parametric statistics (mean ± sample standard deviation) were used for interval and ratio level data while non-parametric statistics (median, range) described nominal and ordinal level data. Variables were further evaluated mathematically and graphically. Mathematical evaluation of sample shape was performed using quintiles, skew, and kurtosis. Box plots, histograms and scatter plots were used to graphically examine groupings and variations within the data. The likelihood ratio chi-squared statistic, Fisher’s two-sided exact test, and Student’s t-test comparison for independent samples were used to assess if there were any significant differences between the cohort and study sample or among important covariates between wear groups.

12.10 Testing the Hypotheses

For the purposes of comparing proportions and for logistic regression, the distribution of the variables was examined and variables were categorized, based on clinically relevant cut points. Although a number of classifications were explored, floor effects and skewed distributions of all pain and function variables resulted in systematically smaller groups with little pain or only some functional difficulties compared to their asymptomatic counter-parts. Methods used to test the specific primary and secondary hypotheses described in detail in Section 11.1 are outlined in this section.

1. To evaluate the relationships between self-reported pain, self-reported function and walking performance, and radiographic markers of potential THR failure (polyethylene wear and periprosthetic boney change), before and after controlling for potential confounders.

   a. Contingency table analysis was performed using the likelihood ratio chi-squared statistic, or Fisher’s exact test in the case of small sample sizes (< 5 per cell), to determine if there was a significant difference in the proportion of patients with at
least some self-reported pain, or self-reported function, as dichotomous variables. The 6MWT pain score compared individuals with no pain to those with some pain (0 versus ≥ 1). The ICOAP score (both subscales and global score) compared individuals with scores indicating only mild if any pain (≤ 25) versus those with more pain (scoring > 25), while the WOMAC-P score compared individuals with mild pain on all questions (scores ≤ 7) versus those with mild to moderate pain (scores > 7). Individuals ‘at-risk’ were defined dichotomously as either those with high annual rates of polyethylene wear (≥ 0.1 mm/year versus < 0.1 mm/year) or at least some periprosthetic boney change (yes/no).

b. The Spearman Rank correlation coefficient was used to test the associations of pain (as both a dichotomous and continuous measure) to radiographic markers of potential THR failure including polyethylene wear (as a continuous variable in mm/year, and as a dichotomous variable) and periprosthetic boney change (as a dichotomous variable).

c. Spearman (ordinal variables with dichotomous or interval variables) and Pearson (dichotomous versus dichotomous or ratio variables, and ratio versus ratio variables) correlations were performed to evaluate the relationships between explanatory variables (measures of self-reported pain, self-reported function and walking performance) and covariates (BMI, time since surgery, pain catastrophizing and comorbidity). Spearman Rank correlation coefficients were calculated for ordinal level variables, while dichotomous variables were evaluated using Pearson correlation coefficients. Interval and ratio variables were evaluated using the standard parametric Pearson correlation coefficient.

d. The unadjusted and adjusted associations between pain (some/less) and polyethylene wear (dichotomous: yes/no or continuous: mm/year) or periprosthetic boney change (yes/no) were assessed using logistic and exact logistic regression (Norman, 2003).

2. To assess the level of agreement between the two radiographic markers of potential THR failure, polyethylene wear and periprosthetic boney change, when classifying individuals ‘at-risk’ for THR failure.
a. A Fleiss-Cohen (quadratic) weighted kappa coefficient was used to evaluate the agreement.

All data analyses in the present study, with the exception of the sample size calculation for the area under the ROC curve below, were generated using SAS software, Version 9.2 of the SAS System for Windows. Copyright 2002 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

12.11 Sample Size Justification

12.11.1 Logistic Regression

Based on an estimated 30% of participants with at least some pain or functional impairment (variably defined), an alpha value of 0.05, and assuming a binomial distribution for the predictor variable, 99 individuals in the total sample should provide 80.1% actual power to detect a 15% difference between high and low risk groups using logistic regression, 95% of the time. A total sample size of 133 individuals should provide 90.2% power to detect a significant difference of 15% between the high and low risk groups.

12.11.2 Area under the Receiver Operating Characteristic Curve

Calculating the error under the ROC curve with 80% power, an alpha of 0.05, and an estimated 50% of participants with polyethylene wear of ≥ 0.2 mm at 5-7 years, leads to the following observations. Forty-nine patients in each group are needed to detect a 15% difference in the association between wear groups. An expected loss of 25% due to a lack of reliable edge detection requires inflating this number for a net result of 66 per group, or a total sample size of 132. To detect a 10% difference, 111 participants are needed for each group so that a total of 222 patients are required. When accounting for an expected loss of 25% as noted above, the total is 296 patients. (This calculation was provided courtesy of Alex Kiss, PhD.)
Figure 5: Flow diagram of cohort exclusions to get sample.

**Sampling Frame**

N=488
Primary Total Hip Replacement Patients Contacted
Surgery January 2001 - June 2004
Sunnybrook Health Sciences Center and/or Holland Orthopedic and Arthritic Institute
Toronto

- 488 contacted
- 167 (34.2%) unreachable, undecided or awaiting appointment
- 8 deceased (1%)
- 68 excluded (13.9%)
- 72 declined (14.8%)

**Cohort**

N=173
Attended Research Appointment
February 1st 2008 - June 30th 2009

- Exclusions (N=63)
  - 1 BMI ≥100 (outlier)
  - 22 Missing X-Rays
  - 40 Alternative Bearing Surface Implants:
    - 21 All-Polyethylene Cups
    - 11 Metal-on-Metal
    - 8 Radiolucent Ceramic Heads

**Study Sample**

N=110

- Sample Characteristics
  - Primary Diagnosis:
    - 97% Osteoarthritis
    - 3% Other
  - Bearing Surface:
    - 99% Standard Metal on Polyethylene
    - 1% Radiopaque Ceramic Head.
Chapter 5 Results

13 Results

13.1 Cohort and Sample

*Cohort:* As indicated in Figure 5 a total of 173 patients met the inclusion criteria and were able to present themselves for examination at one of the two locations between February 1\(^{st}\) 2008 and June 30\(^{th}\) 2009. Of the 173 patients who attended a research appointment, one patient was excluded as an outlier due to a BMI in excess of 100 kg/m\(^{2}\).

*Identification of Study Sample:* Of the remaining 172 patients, 62 participants were excluded because they had either a non-standard bearing surface implant or an incomplete film history. Of these 62, 22 participants were missing x-rays at one or both of the two time points. Of the 22, 19 were missing a bilateral anterior posterior x-ray while 3 were missing a lateral pelvic x-ray. 40 participants were excluded due to their implant characteristics. Of these 40, 21 participants had all-polyethylene acetabular cups that were cemented in place directly to the bone, leaving the software unable to distinguish a clear border between the implant and the cement with sufficient precision to evaluate penetration. 11 participants had metal-on-metal bearing surfaces where there was no appreciable penetration to be measured, while 8 had ceramic femoral heads that were radio-translucent, preventing the Martell software from distinguishing the femoral head border with sufficient precision to provide an accurate estimate of penetration.

*Study Sample:* As a result, the study sample included 110 individuals.

13.2 Study Sample

*Sample characteristics* (Table IV): The mean age of the 110 participants was 62.8±9.4 with a range of 40-80 years at the time of surgery. Sixty-two percent (62.7%) of the sample were women; most were married (72.7%) and retired (59.1%). Ninety-seven percent (n = 102) of the participants had a primary diagnosis of OA with the remainder diagnosed with OA secondary to avascular necrosis or congenital abnormality such as dysplasia. Participants had a mean BMI of 28.4±5.5 kg/m\(^{2}\). Of the 69 patients with complete co-morbidity data, 43.1% had 3 or more co-morbid conditions. The mean follow-up time was 6.1±0.9 years post surgery. Fifty-one percent
had a left-side hip replacement and 99% had a metal on polyethylene bearing surface. A single patient with a non radio-translucent ceramic femoral head was permitted as Martell wear could be adequately assessed. The study sample and the cohort were compared for differences in demographic and surgical characteristics. No statistically significant differences were observed.

### Table IV: Cohort versus study sample 4-8 years after total hip replacement surgery.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Total Available Cohort (n=172)</th>
<th>Study Sample (n=110)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at surgery - mean±SD (range)</strong></td>
<td>62.6±9.6 (38-80)</td>
<td>62.8±9.4 (40-80)</td>
</tr>
<tr>
<td>Female - n (%)</td>
<td>105 (61.1)</td>
<td>69 (62.7)</td>
</tr>
<tr>
<td>Marital status: Single - n (%)</td>
<td>10 (5.8)</td>
<td>4 (3.6)</td>
</tr>
<tr>
<td>Married/common law</td>
<td>128 (74.4)</td>
<td>80 (72.7)</td>
</tr>
<tr>
<td>Widowed</td>
<td>20 (11.6)</td>
<td>16 (14.6)</td>
</tr>
<tr>
<td>Divorced</td>
<td>12 (7.0)</td>
<td>9 (8.2)</td>
</tr>
<tr>
<td>Employment status: Employed - n (%)</td>
<td>56 (32.6)</td>
<td>40 (36.4)</td>
</tr>
<tr>
<td>Retired</td>
<td>110 (64.0)</td>
<td>65 (59.1)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>5 (2.9)</td>
<td>4 (3.6)</td>
</tr>
<tr>
<td>BMI at surgery (kg/m²) - mean±SD (range, n= )</td>
<td>28.9±5.5 (18.1-49.7, n=169)</td>
<td>28.4±5.5 (18.1-46.5, n=109)</td>
</tr>
<tr>
<td>Co-morbidities: 0 or 1 – n/69 (%)</td>
<td>26 (37.7)</td>
<td>17 (33.3)</td>
</tr>
<tr>
<td>2</td>
<td>17 (24.6)</td>
<td>12 (23.5)</td>
</tr>
<tr>
<td>3+</td>
<td>26 (37.7)</td>
<td>22 (43.1)</td>
</tr>
<tr>
<td>Pain catastrophizing scale (/52) - median (range)</td>
<td>15 (0-52)</td>
<td>15.5 (0-52)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surgical characteristics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary diagnosis: Osteoarthritis - n (%)</td>
<td>163 (94.8)</td>
<td>102 (92.7)</td>
</tr>
<tr>
<td>Other¹</td>
<td>9 (5.2)</td>
<td>8 (7.3)</td>
</tr>
<tr>
<td>Years post surgery - mean±SD (range)</td>
<td>6.2±0.9 (4.4-7.9)</td>
<td>6.1±0.9 (4.4-8.0)</td>
</tr>
<tr>
<td>Replacement hip: Left - n (%)</td>
<td>89 (51.7)</td>
<td>57 (51.8)</td>
</tr>
</tbody>
</table>

¹ Avascular Necrosis, Congenital Hip Dysplasia, Protrusio Acetabuli, other

P value not significant for all comparisons at P < 0.05.
13.3 Explanatory variables

13.3.1 Measures of Self-Reported Pain Post THR

There were few participants with significant pain or disability as indicated in Table V, found after Figure 6 below. All measures had a median between 0 and 1. The 6MWT pain median score was 0 with a range of 0-9 out of 10. Similar trends were seen with the ICOAP: constant pain subscale (median 0: 0-75/100), intermittent pain subscale (median 0: 0-62.5/100), total score (median 0: 0-59.1/100); and the WOMAC pain subscale (median 0: 0-14/20).

Average arthritis pain intensity had a median score of 0.55 (0-9.7/10). Most participants (71.8%) did not use any hip pain medication and only 11.8% of participants used mobility aids of any kind. Pain scores on all measures, with the exception of the Numeric Rating Scale (NRS) for Arthritis Pain, showed a strong floor effect with a median score of 0, but a wide range of values. The distributions of the respective pain scores were extremely left-skewed with the majority of scores falling in the first third of the scale as indicated in Figure 6. No more than a 1/3 of patients reported at least some pain on any measure.

Six minute walk test pain scores were based on 101 individuals (91% of the possible sample) as nine patients either refused due to time restrictions or were unable to complete the six minute walk test or the pain scale at the study appointment. Only 30.7% of the analysis sample fell into the a priori defined higher pain group (score ≥1). Approximately eight percent of participants had significant ICOAP constant pain (>25/100 as defined a priori) while 9.2% had some intermittent pain (>25/100). Only seven percent of participants had more than mostly ‘mild’ WOMAC pain (>7/20). The Numeric Rating Scale (NRS) for arthritis pain was the only exception to the trends noted above as it had a median of 0.55 with scores on both ends of the scale. Unfortunately, wording of the measure was not hip-specific so that it was not expected to exhibit hip-specific relationships. As a result, it was dropped from further consideration.

The median pain catastrophizing score was 15.5 (min-max = 0-52/52). Spearman Rank correlations between pain catastrophizing scores and all pain measures were low (ρ < 0.2).
Figure 6: Histograms of self-reported pain scores. A) Six minute walk test pain  B) ICOAP Intermittent subscale score normalized /100  C) WOMAC pain subscale /20.
Table V: Characteristics of study sample 4-8 years after total hip replacement surgery (n=110).

<table>
<thead>
<tr>
<th>Arthritis Severity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility aids: None - n (%)</td>
<td>97 (88.2)</td>
</tr>
<tr>
<td>One cane</td>
<td>6 (5.5)</td>
</tr>
<tr>
<td>Two canes</td>
<td>1 (0.91)</td>
</tr>
<tr>
<td>Walker</td>
<td>5 (4.6)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.91)</td>
</tr>
<tr>
<td>Pain medication: No - n (%)</td>
<td>79 (71.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>25 (22.7)</td>
</tr>
<tr>
<td>Don't know</td>
<td>3 (2.7)</td>
</tr>
</tbody>
</table>

Self-Reported Pain

<table>
<thead>
<tr>
<th>Description</th>
<th>Median (Range, n=)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six minute walk test (6MWT) pain score (/10)</td>
<td>0 (0-9, n=101)</td>
</tr>
<tr>
<td>ICOAP constant pain subscale (/100)</td>
<td>0 (0-75)</td>
</tr>
<tr>
<td>ICOAP intermittent pain subscale (/100)</td>
<td>0 (0-62.5, n=109)</td>
</tr>
<tr>
<td>ICOAP constant and intermittent pain (/100)</td>
<td>0 (0-59.1, n=109)</td>
</tr>
<tr>
<td>WOMAC pain subscale (/20)</td>
<td>0 (0-14, n=109)</td>
</tr>
<tr>
<td>Average arthritis pain intensity (/10)</td>
<td>0.55 (0-9.7)</td>
</tr>
</tbody>
</table>

Self-Reported Function

<table>
<thead>
<tr>
<th>Description</th>
<th>Median (Range, n=)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMAC physical function subscale (/68)</td>
<td>4.5 (0-52, n=109)</td>
</tr>
</tbody>
</table>

Walking Performance

<table>
<thead>
<tr>
<th>Description</th>
<th>Mean±SD (Range, n=)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six minute walk test distance (m)</td>
<td>442.7±140.1 (0-940, n=108)</td>
</tr>
<tr>
<td>Normalized six minute walk test scores</td>
<td>93.9±27.4 (0-207.4, n=108)</td>
</tr>
</tbody>
</table>

1 Normalized to the percentage of the expected value for healthy individuals of equivalent age
13.3.2 Self-Reported Physical Function after THR

The WOMAC functional subscale median value was 4.5 based on 109 participants. Scores ranged from 0 to 52 out of a possible 68. Only 16.5% of individuals had values >17, indicating more than mild difficulty. Similar floor effects were seen with function as were observed with the pain measures (Figure 7).

Figure 7: Histogram of self-reported function scores.
13.3.3 Walking Performance after THR

Six minute walk test performance values were significantly less skewed (Figure 8) compared to self-reported pain and self-reported function. Participants walked a mean of 442.7±140.1 m (min-max = 0-940 m), which reflects 93.9±27.4 percent of the average predicted value for a healthy control participant matched in sex, age, weight and height.

Figure 8: Histograms of walking performance scores.
13.3.4 Correlations of Explanatory Variables

The correlations (Table VI) between the five measures of self-reported pain were excellent to poor (<0.20). In Table VI, ρ represents a correlation coefficient while P represents a value of probability.

13.3.4.1 Correlations within Measures of Self-Reported Pain

The 6MWT pain score showed fair to moderate correlations with all other measures of self-reported pain (ρ=0.36-0.42) and was most highly correlated with the ICOAP total score. The ICOAP total score showed almost perfect correlation with its intermittent subscale and was highly correlated to the WOMAC pain subscale (ρ=0.84). The ICOAP intermittent subscale showed good correlation to the WOMAC pain subscale (ρ=0.75) and moderate correlations (<0.53) with all other self-report measures of pain, except 6MWT pain with which it was moderately correlated (ρ=0.40). The ICOAP constant pain subscale showed fair to moderate correlations with other measures of self-reported pain (ρ=0.38-0.52). The WOMAC pain subscale correlated well with all other measures of self-reported pain (ρ=0.52-0.83) except 6MWT pain (ρ=0.36). All confounders and covariates including years after surgery, age at surgery, BMI at surgery and pain catastrophizing score showed poor correlations with all measures of self-reported pain. In all cases correlations were attenuated with dichotomized variables when compared to the ordinal or continuous equivalent.

13.3.4.2 Correlations between Measures of Self-Reported Pain and Self-Reported Function or Walking Performance

Moderate to good correlations were seen between self-reported pain and self-reported function (ρ=0.39-0.63). The highest correlations were seen with the WOMAC pain subscale. Walking performance correlations for distance measurements with self-reported pain were poor (ρ=−0.17 to -0.33) and worse with normalized scores (-0.08 to -0.32). Negative values indicate an inverse relationship between self-reported pain and distance walked or normalized walking performance score, as expected. Attenuation of correlation was again observed when correlations were performed with dichotomized variables.
13.3.4.3 Correlations between Measures of Self-Reported Function and Walking Performance & Covariates

Self-reported function was moderately correlated with functional walking performance as either a distance ($\rho=-0.38$) or normalized ($\rho=-0.33$) measurement. Poor correlations ($\rho<0.20$) with all potential covariates were observed, including years after surgery, age at surgery, BMI at surgery and pain catastrophizing score, with the exception of age at surgery which showed a fair correlation with distance walked ($\rho=-0.39$). As above, dichotomized variables showed attenuation of correlation.
Table VI: Correlation coefficients (Pearson or Spearman Rank) for explanatory variables.

<table>
<thead>
<tr>
<th>Pearson/Spearman p-value</th>
<th>ICOAP Total Pain Score</th>
<th>WOMAC Total Score</th>
<th>Six Minute Walk Test</th>
<th>WOMAC Pain Score</th>
<th>Function</th>
<th>Performance*</th>
<th>Patient Characteristics/Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P-value if &gt; 0.05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Years After Surgery</strong> (decimal)</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.082</td>
<td></td>
<td></td>
<td></td>
<td>0.036</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
<td>0.58</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICOAP Total Pain Score</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.087</td>
<td></td>
<td></td>
<td></td>
<td>0.057</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td>0.084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICOAP Intermittent</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td>Pain Subscale</td>
<td>0.58</td>
<td>0.60</td>
<td>0.53</td>
<td></td>
<td>-0.19</td>
<td>-0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>-0.26</td>
<td>-0.26</td>
<td></td>
<td>0.036</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.004</td>
<td>0.048</td>
<td>0.048</td>
<td></td>
<td>0.053</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td>ICOAP Constant Pain</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td>Subscale</td>
<td>0.75</td>
<td>0.60</td>
<td>0.52</td>
<td></td>
<td>-0.017</td>
<td>-0.062</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.52</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td>-0.12</td>
<td>0.12</td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
<td>0.12</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td>ICOAP Intermittent</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td>Pain Subscale, Bicatoma</td>
<td>0.51</td>
<td>0.55</td>
<td>0.42</td>
<td></td>
<td>-0.19</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>-0.26</td>
<td>-0.26</td>
<td></td>
<td>0.080</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td>0.12</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td>ICOAP Constant Pain</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td>Subscale</td>
<td>0.51</td>
<td>0.55</td>
<td>0.42</td>
<td></td>
<td>-0.19</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>-0.26</td>
<td>-0.26</td>
<td></td>
<td>0.080</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td>0.12</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td>WOMAC Pain Subscale</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.51</td>
<td>0.55</td>
<td>0.42</td>
<td></td>
<td>-0.19</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>-0.26</td>
<td>-0.26</td>
<td></td>
<td>0.080</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td>0.12</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td>WOMAC Pain Subscale</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td>Bicatoma</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
<td>-0.19</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.19</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td>0.080</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td>0.12</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td>Six Minute Walk Test</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td>Pain Score</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td></td>
<td>-0.19</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>-0.26</td>
<td>-0.26</td>
<td></td>
<td>0.080</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.71</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
<td>0.12</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
<tr>
<td>Six Minute Walk Test</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td><strong>Age at Surgery</strong></td>
</tr>
<tr>
<td>Pain Score, Bicatoma</td>
<td>0.37</td>
<td>0.37</td>
<td>0.37</td>
<td></td>
<td>-0.14</td>
<td>-0.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>0.17</td>
<td>0.17</td>
<td></td>
<td>-0.026</td>
<td></td>
<td><strong>BMI at Surgery</strong></td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.17</td>
<td>0.17</td>
<td></td>
<td>0.09</td>
<td></td>
<td><strong>Pain Catastrophizing</strong></td>
</tr>
</tbody>
</table>

1 Normalized to the expected value for healthy individuals of equivalent age.
2 Higher scores represent better function as opposed to self-reported function and pain where higher scores indicate more pain or disability.
* P < 0.05,  ** R < 0.001.
Table VI (Cont’d): Correlation coefficients (Pearson or Spearman Rank) for explanatory variables.

<table>
<thead>
<tr>
<th>Pearson/Spearman p-value</th>
<th>Pain</th>
<th>Function</th>
<th>Performance</th>
<th>Patient Characteristics/Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICOAP</td>
<td>WOMAC</td>
<td>Subscale</td>
<td>Years After Surgery (decimal)</td>
</tr>
<tr>
<td></td>
<td>Subscale</td>
<td>Subscale</td>
<td>Subscale</td>
<td>Distance (m)</td>
</tr>
<tr>
<td></td>
<td>Total Score</td>
<td>Intermittent Subscale</td>
<td>Constant Subscale</td>
<td>Pain Subscale</td>
</tr>
<tr>
<td>WOMAC Function Subscale</td>
<td>1.0</td>
<td>0.65 **</td>
<td>-0.38 **</td>
<td>-0.33 **</td>
</tr>
<tr>
<td>WOAMC Function Subscale, Dichotomized (( \leq 17 ))</td>
<td>1.0</td>
<td>-0.33 **</td>
<td>-0.26 *</td>
<td>-0.041</td>
</tr>
<tr>
<td>Six Minute Walk Test Performance Score, Distance (m)</td>
<td>1.0</td>
<td>0.88 **</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>Six Minute Walk Test Performance Score, Normalized Distance</td>
<td>1.0</td>
<td>0.090</td>
<td>0.035</td>
<td>-0.11</td>
</tr>
<tr>
<td>Years After Surgery (decimal)</td>
<td>1.0</td>
<td>0.075</td>
<td>0.44</td>
<td>0.0022</td>
</tr>
<tr>
<td>Age at Surgery</td>
<td>1.0</td>
<td>-0.046</td>
<td>0.63</td>
<td>-0.14</td>
</tr>
<tr>
<td>BMI at Surgery</td>
<td>1.0</td>
<td>0.19</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Pain Catastrophizing</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Normalized to the expected value for healthy individuals of equivalent age.

** Higher scores represent better function as opposed to self-reported function and pain where higher scores indicate more pain or disability.

* \( P < 0.05, \) ** \( R < 0.001. \)
13.4 Outcomes

13.4.1 Measures of Polyethylene Wear

As shown in Table VII, twenty-nine individuals (26.4%) had annual polyethylene wear rates above the osteolysis threshold of 0.1 mm/year, the threshold defined *a priori* to classify patients ‘at-risk’ for THR failure. Eighty-one participants (73% of sample) had annual wear rates < 0.1 mm/year and thus were classified as ‘low-risk’ for THR failure. Mean total linear wear (Table VIII) was 0.44±0.54 mm (0.79-2.71). Some negative values were noted. Mean total volumetric wear was 231.3±258.7 mm³ (range: 1-1417.3). Mean annual volumetric wear was 38.2±40.2 mm³/year (range: 0.18 to 214.7). Mean annual linear wear for the sample was 0.07±0.09 mm/year (range: -0.16 to 0.43 mm/year).

<table>
<thead>
<tr>
<th>Annual wear threshold (mm/year)</th>
<th>&lt; 0.1</th>
<th>0.10-0.19</th>
<th>≥ 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number in analysis subsample - n (%)</td>
<td>81 (73.6)</td>
<td>20 (18.2)</td>
<td>9 (8.2)</td>
</tr>
</tbody>
</table>
Table VIII: Radiographic markers of potential THR failure in study sample 4-8 years after total hip replacement surgery (n=110).

Measures of Polyethylene Wear

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean±SD (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total linear wear (mm)</td>
<td>0.44±0.54 (-0.79-2.71)</td>
</tr>
<tr>
<td>Total volumetric wear (mm³)</td>
<td>231.3±258.7 (1-1417.3)</td>
</tr>
<tr>
<td>Annual linear wear rate (mm/year)</td>
<td>0.07±0.09 (-0.16-0.43)</td>
</tr>
<tr>
<td>Annual volumetric wear rate (mm³/year)</td>
<td>38.2±40.2 (0.18-214.7)</td>
</tr>
</tbody>
</table>

Measures of Periprosthetic Bone Change

<table>
<thead>
<tr>
<th>Measure</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any periprosthetic bone change - n (%)</td>
<td>40 (36.4)</td>
</tr>
<tr>
<td>Any periprosthetic bone change excluding osteopenia - n (%)</td>
<td>33 (30.0)</td>
</tr>
<tr>
<td>Any osteopenia 2˚ to stress shielding - n (%)</td>
<td>10 (9.1)</td>
</tr>
<tr>
<td>Any osteolysis - n (%)</td>
<td>6 (5.5)</td>
</tr>
<tr>
<td>Any cortical hypertrophy - n (%)</td>
<td>6 (5.5)</td>
</tr>
<tr>
<td>Any pedestal - n (%)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Any heterotrophic ossification - n (%)</td>
<td>24 (21.8)</td>
</tr>
</tbody>
</table>

13.4.2 Measures of Periprosthetic Bone Change

As indicated in Table VIII, thirty-six percent of patients (n=40) had some periprosthetic bone change, defined as any periprosthetic bone change (osteolysis, osteopenia, cortical hypertrophy, pedestal or heterotopic ossification) in any zone; 33 patients met our second definition of any periprosthetic bone change which excluded measures of osteopenia indicating stress shielding (osteolysis, cortical hypertrophy, pedestal or heterotopic ossification), and 24 patients met our third definition of any periprosthetic bone change (presence of any heterotopic ossification - of any Brooker’s grade).

13.5 Agreement of Polyethylene Wear & Periprosthetic Bone Change When Classifying Individuals ‘At-Risk’ for THR Failure

Agreement, adjusted for chance (Table IX), between polyethylene wear, evaluated using annual linear polyethylene wear (≤0.1 mm/year versus >0.1 mm/year), and any periprosthetic bone change (yes/no) in any zone, defined three ways, was poor (kappa = 0.03-0.10).
Table IX: Cross-tabular frequencies and agreement statistics between potential radiographic markers of total hip replacement failure.

A. Annual linear wear versus periprosthetic boney change (PPBC)

<table>
<thead>
<tr>
<th></th>
<th>Any PPBC</th>
<th>No PPBC</th>
<th>Row total</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Row percent</td>
<td>Col percent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥0.1 mm/yr</td>
<td>12</td>
<td>17</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.9</td>
<td>15.5</td>
<td>26.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41.4</td>
<td>58.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30.0</td>
<td>24.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.1 mm/yr</td>
<td>28</td>
<td>53</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.5</td>
<td>48.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34.6</td>
<td>65.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70.0</td>
<td>75.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col total</td>
<td>40</td>
<td>70</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Col percent</td>
<td>36.4</td>
<td>63.6</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Quadratic weighted kappa = 0.0607

B. Annual linear wear versus PPBC excluding osteopenia

<table>
<thead>
<tr>
<th></th>
<th>Any PPBC excluding osteopenia</th>
<th>No PPBC excluding osteopenia</th>
<th>Row total</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Row percent</td>
<td>Col percent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥0.1 mm/yr</td>
<td>11</td>
<td>18</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.0</td>
<td>16.4</td>
<td>26.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37.9</td>
<td>62.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>33.3</td>
<td>23.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.1 mm/yr</td>
<td>22</td>
<td>59</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20.0</td>
<td>53.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27.2</td>
<td>72.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>66.7</td>
<td>76.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col total</td>
<td>33</td>
<td>77</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Col percent</td>
<td>30.0</td>
<td>70.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Quadratic weighted kappa = 0.1031

C. Annual linear wear versus heterotopic ossification

<table>
<thead>
<tr>
<th></th>
<th>Any heterotopic ossification</th>
<th>No heterotopic ossification</th>
<th>Row total</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Row percent</td>
<td>Col percent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥0.1 mm/yr</td>
<td>7</td>
<td>22</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.4</td>
<td>20.0</td>
<td>26.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24.1</td>
<td>75.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29.2</td>
<td>25.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.1 mm/yr</td>
<td>17</td>
<td>64</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.5</td>
<td>58.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21.0</td>
<td>79.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70.8</td>
<td>74.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col total</td>
<td>24</td>
<td>86</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Col percent</td>
<td>21.8</td>
<td>78.2</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Quadratic weighted kappa = 0.0333
13.6 Relationship of Radiographic Markers of Potential THR Failure to Self-reported Pain, Self-reported Function & Walking Performance

To assess the associations involving self-reported pain, self-reported function and walking performance, a priori categorizations of all ordinal predictors were used for contingency table analysis (data not shown) followed by logistic regression in order to make results clinically useful. Categories for pain were 6MWT pain: no pain (0) versus any pain (≥ 1), ICOAP scores (constant pain subscale, intermittent pain subscale, and total score): none or only mild pain (≤ 25) versus more pain (> 25), WOMAC pain subscale: none or less than mild pain (≤ 7) versus high mild to moderate pain (≥ 8). Categorizations for self-reported function were none to mild disability on the WOMAC physical subscale (≤ 17) versus moderate or more disability (≥ 18). Walking performance, as both a distance in meters and as a normalized score, was used as a continuous predictor variable.

13.6.1 Relationship of Polyethylene Wear to Self-Reported Pain, Self-Reported Function & Walking Performance

No significant relationships were found (Table X) between annual rates of polyethylene wear and any of our measures of self-reported pain. However, there was a non-significantly higher proportion of patients with ‘at least some pain’ after the six minute walk test in the high annual polyethylene wear group (40.0% versus 27.6%, P=0.32) demonstrating the hypothesized trend, provided an association exists. Moreover, the proportion of patients with ‘at least some’ pain on the ICOAP intermittent pain scale was greater in the high wear versus the low wear group (10.7% versus 8.6% respectively), but this difference was also not statistically significant (P=0.71) (Figure 9).

There was no evidence of a significant association between any measure of function and annual polyethylene wear (Table XI, Figure 10, Figure 11). As bivariate associations were not significant, multiple logistic regression was not performed.
Figure 9: Charts of self-reported pain against annual linear wear. Scores from three measures of pain examined in the study are plotted against annual linear wear as determined by Martell analysis. The dashed lines indicate the maximum values for corresponding pain scores and the range of wear rates observed in the study. The arrows indicate the expected increase in pain score with increasing wear rate for a strong correlation between pain and wear. The sample size varied between 101 and 110 participants.
### Table X: Explanatory variables by wear group.

<table>
<thead>
<tr>
<th></th>
<th>Analysis Subsample (n=110)</th>
<th>Radiographic Wear: No &lt; 0.1 millimetres of penetration/year (n=81)</th>
<th>Radiographic Wear: Yes ≥ 0.1 millimetres of penetration/year (n=29)</th>
<th>P &lt; value (Low Wear versus High Wear)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at surgery - mean±SD (range)</td>
<td>62.8±9.4 (40-80)</td>
<td>62.7±9.4 (40-80)</td>
<td>63.0±9.7 (48-79)</td>
<td>0.88</td>
</tr>
<tr>
<td>Female – n (%)</td>
<td>69 (62.7)</td>
<td>53 (65.4)</td>
<td>16 (55.2)</td>
<td>0.37</td>
</tr>
<tr>
<td>BMI at surgery (kg/m^2) - mean±SD (range, n=)</td>
<td>28.4±5.5 (18.1-46.5, n=109)</td>
<td>28.3±5.6 (20.1-46.5, n=80)</td>
<td>28.8±5.3 (38.8-18.1)</td>
<td>0.66</td>
</tr>
<tr>
<td>Pain catastrophizing scale - median (range)</td>
<td>15.5 (0-52 )</td>
<td>15 (0-49)</td>
<td>11 (0-52)</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>Surgical characteristics and arthritis severity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years post surgery - mean±SD (range)</td>
<td>6.07±0.89 (4.36-7.95 )</td>
<td>6.03±0.91 (4.36-7.95)</td>
<td>6.18±0.86 (5.06-7.85)</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>Pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6MWT pain score:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pain – n (%)</td>
<td>70 (69.3)</td>
<td>55 (72.4)</td>
<td>15 (60.0)</td>
<td></td>
</tr>
<tr>
<td>Any pain – n (%)</td>
<td>31 (30.7)</td>
<td>21 (27.6)</td>
<td>10 (40.0)</td>
<td>0.32</td>
</tr>
<tr>
<td>ICOAP constant pain subscale:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pain – n (%)</td>
<td>101 (91.8)</td>
<td>74 (91.4)</td>
<td>27 (93.1)</td>
<td></td>
</tr>
<tr>
<td>Any pain – n (%)</td>
<td>9 (8.2)</td>
<td>7 (8.6)</td>
<td>2 (6.9)</td>
<td>1.00</td>
</tr>
<tr>
<td>ICOAP intermittent pain subscale:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None/a little pain (≤ 25) – n (%)</td>
<td>99 (90.8)</td>
<td>74 (91.4)</td>
<td>25 (89.3)</td>
<td></td>
</tr>
<tr>
<td>More pain (&gt; 25) – n (%)</td>
<td>10 (9.2)</td>
<td>7 (8.6)</td>
<td>3 (10.7)</td>
<td>0.71</td>
</tr>
<tr>
<td>ICOAP constant and intermittent pain:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None/a little pain (≤ 25) – n (%)</td>
<td>100 (91.7)</td>
<td>74 (91.4)</td>
<td>26 (92.9)</td>
<td></td>
</tr>
<tr>
<td>More pain (&gt; 25) – n (%)</td>
<td>9 (8.3)</td>
<td>7 (8.6)</td>
<td>2 (7.1)</td>
<td>1.00</td>
</tr>
<tr>
<td>WOMAC pain subscale:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None/a little pain (≤ 7) – n (%)</td>
<td>101 (92.7)</td>
<td>73 (91.3)</td>
<td>28 (96.6)</td>
<td></td>
</tr>
<tr>
<td>More pain (&gt; 7) – n (%)</td>
<td>8 (7.3)</td>
<td>7 (8.8)</td>
<td>1 (3.5)</td>
<td>0.68</td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC physical function subscale:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None/a little difficulty (≤ 17) – n (%)</td>
<td>91 (83.5)</td>
<td>65 (81.3)</td>
<td>26 (89.7)</td>
<td></td>
</tr>
<tr>
<td>More difficulty (&gt; 17) – n (%)</td>
<td>18 (16.5)</td>
<td>15 (18.8)</td>
<td>3 (10.3)</td>
<td>0.39</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normalized six minute walk test scores¹ - mean±SD (range, n=)</td>
<td>93.9±27.4 (0-207.4, n=108)</td>
<td>92.3±29.4 (0-207.4, n=80)</td>
<td>98.3±20.2 (45.2-147.4, n=28)</td>
<td>0.24</td>
</tr>
</tbody>
</table>

¹ Normalized to the percentage of the expected value for healthy individuals of equivalent age using Enright's predictive equation
**Table XI:** Univariate logistic regression (dependant variable = wear ≥ 0.1mm/year).

<table>
<thead>
<tr>
<th>Model</th>
<th>Reference</th>
<th>Estimate</th>
<th>Pr &gt; Chi Sq</th>
<th>Odds Ratio</th>
<th>95% Wald CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Patient Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at surgery</td>
<td>per year</td>
<td>0.0037</td>
<td>0.87</td>
<td>1.00</td>
<td>(0.96,1.05)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>0.43</td>
<td>0.33</td>
<td>1.54</td>
<td>(0.65,3.65)</td>
</tr>
<tr>
<td>Years post surgery</td>
<td>per year</td>
<td>0.19</td>
<td>0.43</td>
<td>1.21</td>
<td>(0.75,1.94)</td>
</tr>
<tr>
<td>BMI at surgery</td>
<td>per kg/m²</td>
<td>0.017</td>
<td>0.67</td>
<td>1.02</td>
<td>(0.94,1.10)</td>
</tr>
<tr>
<td>Pain catastrophizing scale</td>
<td>per point</td>
<td>0.00079</td>
<td>0.95</td>
<td>1.00</td>
<td>(0.975,1.03)</td>
</tr>
<tr>
<td><strong>Pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICOAP total score (≤ 25)</td>
<td>None or mild pain</td>
<td>-0.21</td>
<td>0.80</td>
<td>0.81</td>
<td>(0.16, 4.17)</td>
</tr>
<tr>
<td>ICOAP intermittent subscale (≤ 25)</td>
<td>None or mild pain</td>
<td>0.24</td>
<td>0.74</td>
<td>1.27</td>
<td>(0.305, 5.28)</td>
</tr>
<tr>
<td>Pain after six minute walk test (≥ 1)</td>
<td>No pain</td>
<td>0.56</td>
<td>0.25</td>
<td>1.75</td>
<td>(0.68,4.49)</td>
</tr>
<tr>
<td>WOMAC pain subscale (≤ 7)</td>
<td>None or mild pain</td>
<td>-0.99</td>
<td>0.37</td>
<td>0.37</td>
<td>(0.044,3.17)</td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC physical function</td>
<td>None or mild difficulty</td>
<td>-0.69</td>
<td>0.30</td>
<td>0.50</td>
<td>(0.13, 1.87)</td>
</tr>
<tr>
<td>subscale (≤ 17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Performance Measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Six minute walk test scores (m)</td>
<td>per metre</td>
<td>0.0013</td>
<td>0.44</td>
<td>1.00</td>
<td>(0.10, 1.00)</td>
</tr>
<tr>
<td>Six minute walk test scores, normalized¹</td>
<td>per percent predicted value</td>
<td>0.0083</td>
<td>0.32</td>
<td>1.01</td>
<td>(0.99,1.03)</td>
</tr>
</tbody>
</table>

¹ Normalized to the percentage of the expected value for healthy individuals of equivalent age.
Figure 10: Chart of self-reported function against annual linear wear. Scores of WOMAC Physical Function are plotted against annual linear wear as determined by Martell analysis. The dashed lines indicate maximum values for the function subscale and the range of wear rates observed in the study. The arrow indicates the expected increase in function score with increasing wear rate for a strong correlation between the two variables. The sample size was 109 patients.
Figure 11: Chart of walking performance against annual linear wear. The distance in metres walked by participants during the 6-Minute Walk Test is plotted against annual linear wear as determined by Martell analysis. The dashed lines indicate the upper limits for corresponding variables as observed in the study while the arrow indicates the expected decrease in walk test score with increasing wear. The maximum distance walked was 940 metres and the sample size was 101 individuals.
13.7 Relationship of Periprosthetic Bone Change to Self-Reported Pain, Self-reported Function & Walking Performance

No significant relationships were found with periprosthetic bone change and any of our measures of self-reported pain, self-reported function or walking performance regardless of the definition of periprosthetic change used (Table XII).

13.8 Relationship of Radiographic Markers of Potential THR Failure to Covariates

No significant relationships were found with any radiographic marker of potential THR failure and various patient characteristics that were suspected confounders or effect modifiers (Table XI, Table XII).
Table XII: Univariate logistic regression (dependant variable = periprosthetic boney change (yes/no)).

<table>
<thead>
<tr>
<th>Model</th>
<th>Reference</th>
<th>Estimate</th>
<th>Pr &gt; Chi Sq</th>
<th>Odds Ratio</th>
<th>95% Wald CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at surgery</td>
<td>per year</td>
<td>-0.01</td>
<td>0.53</td>
<td>0.99</td>
<td>(0.95, 1.03)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>0.51</td>
<td>0.21</td>
<td>1.67</td>
<td>(0.75, 3.71)</td>
</tr>
<tr>
<td>Years post surgery</td>
<td>per year</td>
<td>0.058</td>
<td>0.80</td>
<td>1.06</td>
<td>(0.685, 1.62)</td>
</tr>
<tr>
<td>BMI at surgery</td>
<td>per kg/m(^2)</td>
<td>-0.005</td>
<td>0.89</td>
<td>0.995</td>
<td>(0.93, 1.07)</td>
</tr>
<tr>
<td>Pain catastrophizing scale</td>
<td>per point</td>
<td>-0.002</td>
<td>0.85</td>
<td>1.00</td>
<td>(0.97, 1.00)</td>
</tr>
<tr>
<td>Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICOAP total score (≤ 25)</td>
<td>None or mild pain</td>
<td>0.35</td>
<td>0.62</td>
<td>1.42</td>
<td>(0.36, 5.64)</td>
</tr>
<tr>
<td>ICOAP intermittent subscale (≤ 25)</td>
<td>None or mild pain</td>
<td>0.15</td>
<td>0.82</td>
<td>1.17</td>
<td>(0.31, 4.41)</td>
</tr>
<tr>
<td>Pain after six minute walk test (≥ 1)</td>
<td>No pain</td>
<td>-0.22</td>
<td>0.64</td>
<td>0.81</td>
<td>(0.33, 1.97)</td>
</tr>
<tr>
<td>WOMAC pain subscale (≤ 7)</td>
<td>None or mild pain</td>
<td>-0.55</td>
<td>0.51</td>
<td>0.58</td>
<td>(0.11, 3.01)</td>
</tr>
<tr>
<td>Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC physical function subscale (≤ 17)</td>
<td>None or mild difficulty</td>
<td>-0.87</td>
<td>0.12</td>
<td>0.42</td>
<td>(0.13, 1.87)</td>
</tr>
<tr>
<td>Performance Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Six minute walk test scores (m)</td>
<td>per metre</td>
<td>0.001</td>
<td>0.63</td>
<td>1.00</td>
<td>(0.998, 1.004)</td>
</tr>
<tr>
<td>Six minute walk test scores, normalized(^1)</td>
<td>per percent predicted value</td>
<td>-0.001</td>
<td>0.84</td>
<td>1.00</td>
<td>(0.98, 1.01)</td>
</tr>
</tbody>
</table>

\(^1\) Normalized to the percentage of the expected value for healthy individuals of equivalent age
Chapter 6 Discussion

14 Purpose

In this chapter, the results described in Chapter 5 are discussed in relation to the objectives outlined in Chapter 3, in the context of the existing literature. Conclusions are drawn, the impact of the data structure is examined, and the advantages and limitations of the design and implementation of the present study are discussed. Suggestions for future work are presented in Chapter 7.

15 Overview and Study Objectives

In the review of prior work, no studies were identified as having, as their primary objective, comprehensively evaluated the relationship of self-reported pain and function, or measures of walking performance, to radiographic markers of potential THR failure using newer technologies and recent advances in measurement techniques. As current methods of in-person surveillance are unsustainable in the long-term, such evaluations are necessary to identify a viable and cost-effective alternative. The objective of the present study was to address this gap. Specifically, investigators sought to answer three questions: (1) Could a candidate patient-centered screening tool be identified? (2) Do self-reported pain and function or walking performance relate to radiographic markers of potential THR failure (polyethylene wear and periprosthetic boney change) 4 to 8 years post-THR surgery due to OA? (3) Do polyethylene wear and periprosthetic boney change identify the same individuals as ‘at-risk’ for THR failure?

16 Main Results Summarized

1) A candidate screening tool could not be identified from the evidence accumulated.

2) There was no conclusive evidence of a cross-sectional association of self-reported measures of pain to radiographic markers of potential THR failure (polyethylene wear or periprosthetic boney change).

3) A greater proportion of patients with pain after the six minute walk test and/or intermittent pain had significant (greater than 0.1 mm/year) polyethylene wear, but these potential associations were not statistically significant. This is the first study to
highlight the importance of pain in future studies leading to development of a post-
operative surveillance tool.

4) There was no evidence of a cross-sectional association of self-reported or objective
measures of function to radiographic markers of potential THR failure (polyethylene
wear or periprosthetic boney change).

5) Power was limited as:

a. Few patients had any significant self-reported pain, or limitations in self-
   reported function or walking performance, 4-8 years post-THR surgery for
   OA.

b. Few patients showed any evidence of radiographic markers of THR failure
   such as a high annual rate of polyethylene wear (≥ 0.1 mm/year) or any
   periprosthetic boney change (yes/no and variably defined) 4-8 years post-THR
   surgery for OA.

6) There is poor agreement between annual linear wear and PPBC, the two radiographic
markers of potential THR failure, when classifying individuals as ‘at-risk’ of THR
failure.

17 Potential Explanations for Negative Findings

There are several possible explanations for the results. These include:

1) There is no association between patient-centered measures and radiographic markers
of potential THR failure,

2) There was insufficient power in this sample to detect an association between patient-
centered measures and radiographic markers of potential THR failure 4-8 years post-THR
surgery due to OA, and

3) Design limitations and bias reduced the power of the study and obscured any evidence
of a possible relationship.
Several factors limited the statistical power to detect significant associations between patient-centered measures and radiographic markers of potential THR failure. These include: (a) the small total sample size (n=110), (b) an insufficient number of patients with significant self-reported pain or functional (self-reported or walking) impairment, and (c) few individuals with significant radiographic markers of potential THR failure (polyethylene wear ≥ 0.1 mm/year or any periprosthetic boney change).

While sample sizes of 100 are generally considered reasonable for a cohort study relying on the collection of clinical information and x-ray data, and smaller samples are frequently employed (Callaghan et al., 2006; Giannikas et al., 2002; Hallan, Lie, & Havelin, 2006; Howie et al., 2007; Keener et al., 2003; Lautiainen et al., 1994; Palm et al., 2007; Strom et al., 2006), in the presence of large variations within the wear measurements, more individuals (>>100) are required to distinguish an effect, if present. As both the wear measurements and the pain and functional evaluations in this study showed considerable variability, it is probable, in retrospect, that the total sample size was insufficient. Forewarned of such limitations, future researchers should aim to assess samples in excess of 200 and/or to make sample size calculations taking into account floor effects. Concurrently, researchers should seek to reduce the variability in the techniques employed by administering questionnaires using a single method and using three time-points to evaluate polyethylene wear instead of two, as discussed below.

In addition to the small sample size, the presence of few patients with significant pain or functional impairment also tended to obscure any positive results (Smythe & Bogoch, 2008). This factor also limited the statistical power of the study, particularly when scores were categorized for the purpose of logistic regression, as this forced categorization at values that are not necessarily clinically relevant. Since good clinical results were expected 4-8 years after surgery (Huo et al., 2007; Huo, Parvizi, Bal, & Mont, 2009), the results in the present work were not particularly surprising, although the small number of patients with any significant impairment was not expected. However, given the improvements in implant technology which delay clinical failure, it is likely that the evolution of radiographic markers is similarly delayed.

Power was further reduced by the rarity of radiographic markers of potential THR failure. Although mean annual wear rates were in line with values noted in comparable samples using the Martell method (Mayman, Anderson, Su, & Sculco, 2007), 80 percent of participants had low
annual wear rates, below the osteolysis threshold of 0.1 mm/year, which resulted in a small symptomatic group and made it difficult to apply logistic regression techniques. Explanations include a) the differences in polyethylene implant liners, and b) the short follow-up period after surgery in the context of the improved implant survival rates noted above. Liners with different compositions and treatment properties were part of the sample and these properties are known to affect rates of polyethylene wear in both conventional and highly cross-linked polyethylene (R. Barrack et al., 2006). While among conventional polyethylene liners, little variation in penetration rate is expected, conventional polyethylene liners can have 10 times the amount of polyethylene wear compared to their highly cross-linked counterparts which made up a small subsample of the cohort (McCalden et al., 2009; Triclot, Grosjean, El Masri, Courpied, & Hamadouche, 2007).

Another explanation for the small proportion of patients with clinically significant radiographic markers of potential THR failure could be that the 4-8 year follow-up period may not be sufficiently long to allow for the development of significant polyethylene wear. With ongoing improvements in materials science and polymer preparation designed to reduce bearing surface wear, longer periods of follow-up are becoming increasingly necessary to show clinically significant annual wear rates. If structural changes are primarily the result of PE debris secondary to wear, then structural changes would most likely occur later in the failure process. In the early stages, structural changes would either be absent or would not yet have progressed to the point of being clinically relevant or detectable. This view is supported by the lack of structural changes observed in many other early follow-up studies (Heilpern, Shah, & Fordyce, 2008; Y. H. Kim, 2008; Rajadhyaksha et al., 2009). Such observations, while inconclusive due to study limitations, may reinforce the expert opinion that the majority of these visits may be a waste of health care resources (Kreder, 2010), and could suggest, along with other literature (Ghoz & Macdonald, 2008), that there is potentially no requirement for radiographic surveillance, in the absence of problems, between years 1-5 and potentially up to 8 years post-surgery for otherwise healthy patients. If implemented, this change in practice could significantly reduce health care spending (Bolz et al., 2010).

Despite current efforts to enhance the clarity of x-ray images, and to mechanize and standardize the evaluation of images, there remains a certain level of subjectivity in the process of evaluating radiographic markers of potential THR failure and, in particular, periprosthetic
boney change. This subjectivity could introduce variability and obscure the presence of an association as even skilled practitioners may differ, to some extent, in their interpretation of the same radiograph (Clohisy et al., 2009). Although few studies exist on the impact of image quality, the generally perceived reduction in the reliability of computerized programs in reading images of poor quality, suggests that this might be a factor in expert analyses of periprosthetic boney change. Variation is introduced with differing hip anatomy, the exposure angle, and the imaging techniques used to collect the images, which can vary between images and by facility. In spite of efforts to control for some sources of variation, including evaluator training, utilizing surgical professionals for x-ray evaluation, and a consensus process; radiographic evaluation of periprosthetic boney change remains among the most inconsistent of procedures in comparison with other standardized measures (Clohisy et al., 2009).

Moreover, precision component analysis has revealed variations as high as 0.242 mm with the Martell and other computer-based methods of polyethylene wear measurement. This casts doubt on the reliability of this method at very low values of wear (Geerdink et al., 2008). Radiostereometric Analysis (RSA) is a technique for assessing polyethylene wear, with an accuracy of 0.15 mm, which makes use of metal beads implanted in the polyethylene liner in order to get wear measurements independent of bedding-in effects (permanent deformations in the polyethylene liner that do not release wear debris). It has been suggested that RSA is the only method suitable for detecting clinically relevant wear with a precision of 0.2 mm or less.

Unfortunately, RSA requires a prospective design as it requires the placement of the metal beads in the liner at the time of surgery, and this represents a substantial financial burden, making it impractical for use in larger studies and clinical practice (Bragdon et al., 2006; McCalden et al., 2005). Moreover, RSA does not account for creep (displacement of the liner within the metal cup) and, as a result, will likely not be worth the increased cost (Digas, Karrholm, Thanner, Malchau, & Herbergs, 2004; Jacobs, Christensen, Greenwald, & McKellop, 2007a; Jacobs, Christensen, Greenwald, & McKellop, 2007b). In the present work, a number of patients with negative wear values were noted. In view of the wear measurement limitations discussed above, it is assumed that these negative values result from the large measurement errors associated with small changes in wear. This assumption is consistent with previous studies (Bragdon et al., 2007; Martell & Berdia, 1997; McCalden et al., 2009).
Less than 40% of the sample showed evidence of PPBC, which is consistent with previous observations (Huo et al., 2009; Marshall et al., 2008; Giannikas et al., 2002; Vervest et al., 2005). This factor also limited the statistical power of the study making it impossible to evaluate each type of boney change independently. Moreover, of those individuals with some boney change, the majority had only heterotopic ossification, which is generally, except in severe cases, clinically asymptomatic and has no significant relationship to THR failure (Kocic et al., 2010). As noted in the introduction, future investigators would ideally have sufficient statistical power to permit the assessment of each measure of periprosthetic boney change separately, in order to avoid any bias resulting from clinically insignificant measures of boney change not strongly related to future THR failure.

As statistical analyses of these results included multiple tests on a single data set, the issue of adjustment for multiple testing must be considered. While adjusting the threshold probability value to obtain a significant result was previously a preferred method when looking at post hoc multiple comparisons in linear models using methods such as the Bonferroni correction (dividing the acceptable significance threshold, the P-value, by the number of tests), controversy exists as to the requirement for adjustment in the context of multiple testing (Bland & Altman, 1995; Feise, 2002; Goodman, 1998; Greenland & Robins, 1991; Rothman, 1990). Alternatives to P-value adjustment, such as selecting a primary outcome measure without adjustment, is recommended under certain circumstances, such as with multiple outcome measures in clinical trials (Feise, 2002). In the context of the present cohort study, adjustment using the Bonferroni method might account for potential false positive results, by chance, when doing a large number of comparisons but, in the case of non-significant results, is irrelevant as results would remain non-significant. In a broader context, using what is considered to be a very stringent method of adjustment increases the chances of false negatives and leads to a publication bias and a reduction in power (Feise, 2002; Nakagawa, 2004). Future investigators should address the issue of multiple testing, but may want to employ alternatives to a Bonferroni correction.

18 Study Advantages & Moving Forward

The present study incorporated several advances with respect to previous work. First and foremost, a newer, more sensitive measure of pain, the ICOAP, which may be more sensitive to
the hypothesized relationship, was used. This measure permitted investigators to look separately at the two dimensions of pain, intermittent and constant, which are important to OA sufferers. Secondly, this study also investigated associations to THR failure of two measures which are new in this area; pain after a walking performance measure and, walking performance itself. Thirdly, polyethylene wear was evaluated using more advanced technology which was potentially more sensitive to early radiographic change. Finally, several potential confounders and effect modifiers were examined, such as time from surgery, age at surgery, BMI at surgery, comorbidity and the newly identified construct of pain catastrophizing which had not been evaluated in previous work. The present work is also, to the best of our knowledge, the only investigation to date which has examined the relationships between a number of different, technologically advanced measures of self-reported pain and polyethylene wear in a single sample.

Of interest, although not statistically significant, was the suggestion that a higher proportion of individuals with significant annual polyethylene wear was seen among those who had some pain after walking or at least mild pain on the intermittent ICOAP pain scale. While an association between intermittent pain and radiographic markers of potential THR failure (particularly polyethylene wear) was hypothesized from previous work, none of the studies evaluated in the systematic review examined the association between pain after the six minute walk test and failure or its markers. Even though that performance measure is often cited in the current orthopaedic literature, the significance of walk test pain is seldom examined, so that the indication, in the present work, of a potential association between pain after the six minute walk test and radiographic measures of potential failure was both interesting and somewhat surprising. Of note, pain after the six minute walk test was the least affected by the noted floor effects, with 30% of the sample reporting at least some pain and thus classified as ‘at-risk’ for THR failure.

Future research in this area should include these previously neglected or unavailable OA-specific measures of pain, such as pain after an objective performance measure and intermittent pain that may be more relevant to THR failure. In addition, since one suggested mechanism for pain due to polyethylene wear is a localized inflammatory response around the prosthesis due to polyethylene wear debris (Schmalzried & Callaghan, 1999), future investigators may want to assess localized inflammatory markers, provided current limitations in target identification
confounded by other systemic inflammatory diseases can be managed (Bauer, Shanbhag, & Implant Wear Symposium 2007 Biologic Work Group, 2008).

Another strength of this work is the evaluation of new (pain catastrophizing) and previously identified (age, sex, BMI) covariates. No previous studies evaluated coping behaviours, such as pain catastrophizing, in the context of examining the relationship between pain and THR failure. But pain catastrophizing has been identified as a factor affecting the results of self-reported pain in the osteoarthritis population because, in the presence of low scores, high pain catastrophizing can cause an over-representation of pain and spuriously indicate a relationship where one may not be present. As postulated, pain catastrophizing was found over the full range of scores within the present sample, implying that it may have a significant influence on the pain-failure relationship and thus may be an important covariate in future cohort studies.

Technologically more advanced methods of measuring polyethylene wear were also used in the present work. Martell methods have shown better precision and accuracy when compared to manual methods (Martell, Berkson, Berger, & Jacobs, 2003). The present study used a two-dimensional wear vector to provide wear estimates. While it has been suggested that three-dimensional vectors are preferable, they are costly to generate and have reduced reliability compared to their two-dimensional counter-parts (Bragdon et al., 2005; Martell et al., 2003). Moreover, 2-dimensional wear vector annual wear rates are frequently used when discussing osteolysis thresholds, as discussed above. As thresholds (> 0.1 mm/year) were used to classify ‘at-risk’ patients in the present investigation, there is justification for choosing a two-dimensional over a three-dimensional wear vector approach (Dumbleton et al., 2002; Harris, 2003; Sochart, 1999). Given the relative cost to benefit ratio, and in the context of future investigations, the choice of a two-dimensional vector to calculate wear provides a better estimate, over manual techniques, without a significant additional demand for resources and should be considered as the preferred expression of Martell wear rates.

Finally, investigators considered two radiographic markers of potential THR failure to examine the research question, polyethylene wear and periprosthetic boney change. As was also found in previous reports (Ghoz & Macdonald, 2008; Toms et al., 2009), there was poor agreement between the two radiographic markers of potential THR failure. This confirms that,
while there is a link between osteolysis and polyethylene wear, other mechanisms such as stress shielding, inappropriate load distribution from bone to implant, or mechanical micro-motion (possibly but not necessarily related to osteolysis), may cause osteopenia, cortical hypertrophy, and pedestal or contribute to osteolysis. Therefore, measuring these parameters may not directly indicate failure and thus have an unknown impact on evaluating their relationship with patient-centered outcomes. It is possible that polyethylene wear is of greater importance in earlier years while some components of periprosthetic boney change may be more important later on as gross changes become more apparent. Ultimately, since these changes remain rare at 4-8 years post-surgery, in both our sample and others (Cruz-Pardos & Garcia-Cimbrelo, 2001; Spicer et al., 2001), their clinical utility in early stages seems questionable. Radiographic surveillance is possibly unnecessary in the early post-acute follow-up period if patients are asymptomatic. Nevertheless, the clinical importance of these measures, after longer follow-up periods, is established and thus periprosthetic boney change should continue to be evaluated in future research studies looking at longer-term associations of pain, function and performance to radiographic markers of THR failure.

19 Study Limitations

While the advantages of the present study include the nature and variety of patient-centered measures that are more specific, the technologically more advanced evaluation of the primary outcome (polyethylene wear), and control for key patient-centered factors, several limitations must be acknowledged. The study was nested within a larger prospective cohort study which restricted the available sample, only a single academic center was considered, there was variation in the methodology for questionnaire administration, reliable comorbidity was not available for the entire sample and therefore could not be considered, using Martell analysis restricted relevance of results to patients with standard bearing surface implants, costs for Martell analysis of polyethylene wear meant that only 2 time points could be used to estimate annual wear, and the study design was cross-sectional.

This study was nested within a larger prospective study; thus, our study participants were recruited from among those eligible for the larger study and this may have introduced sampling bias. Although ideally, this study would have included individuals with all ranges of pain and wear, the requirement to have younger, uncomplicated patients in the larger study prevented their
inclusion in the present study. Patients who were older (≥ 80) at the time of their THR operation were excluded. As older patients are, in general, less active and may not need the more robust highly cross-linked polyethylene liner that have lower wear values, this may have excluded a group of individuals more likely to have greater wear and/or boney change. Finally, the larger study excluded patients who had hip-replacements due to trauma or other diseases affecting physical function. If these same conditions are associated with increased wear and/or higher risk for THR failure, this may have further reduced the numbers with substantial wear in the smaller study, and thus affected the statistical power.

The sample was drawn from a single, academic hospital in the urban center of Toronto, possibly introducing selection bias as all operating surgeons were high volume, lower extremity joint replacement specialists at the time of THR surgery. These specialists have been shown to have better patient-centered outcomes and lower complication rates when compared to their low volume counter-parts (Bordini et al., 2007), although this is controversial (Gandhi, Tso, Davis, & Mahomed, 2009). As the time dependency of these ‘worse outcomes’ remains unknown, future studies may want to consider that in a sample which includes rural or community hospitals, results may show an earlier or larger association of patient-centered measures such as pain, function or walking performance to wear or other markers of THR failure. Variation in the mode of administering questionnaires may also have added to the significant variation and power limitations discussed above. While there is evidence to support the use of mixed modes as acceptable, since the WOMAC 3.0 (Bellamy, Campbell, Hill, & Band, 2002) and the ICOAP (G. Hawker, 2007; G.A. Hawker, French et al., 2008) are designed as self-administered questionnaires, and have been validated for interviewer administration over the telephone, it is not customary to vary modes within a single study. A significant body of work, cautioning against mixed modes due to a potential negative impact on data quality, also exists (Bowling, 2005). Bowling argues that while different modes are acceptable for a single questionnaire, the potential for bias and variation is greatest ‘between modes’ versus ‘within a given mode’. The amount of bias and the response rates generated vary, even within a single mode, making the impact of mixed modes applied within a single study impossible to predict.

Unfortunately, reliable data on comorbidity was only available for 69 participants due to a switch of comorbidity questionnaires within the larger study from which the present sample was drawn. This meant that investigators were unable to consider this variable as a covariate in
the present analysis. Nevertheless, the previously noted potential impact of comorbidity on pain and function indicate that this covariate should be considered in future studies in order to evaluate the potential effect on any demonstrated association with radiographic markers of potential THR failure.

Technological limitations of the Martell method required the exclusion of patients with alternative bearing surfaces (Martell et al., 2003). This restricted the relevance of our findings to those with standard bearing surfaces and eliminated, to some extent, younger and more active patients for whom alternative implants are traditionally recommended. Since younger patients tend to have worse outcomes (Delaunay, Bonnomet, Clavert, Laffargue, & Migaud, 2008), there is the possibility that this restriction introduced bias in the present work. However, in the context of emerging concerns with respect to the production of metallic debris or the shattering of ceramic bearing surfaces, and the excellent results with highly cross-linked polyethylene, this limitation was deemed justifiable. The possible increase in the use of highly cross-linked polyethylene bearing surfaces may mean that the methods used in the present work will be largely applicable in future studies.

In addition to the limitations imposed by selecting the Martell method, the method of defining the wear variable was also controversial, and may have increased the variability and limited the power of the sample. Two options were available to obtain estimates for polyethylene wear: 1) using two time points to look at the difference between the two, or 2) using three time points and comparing slopes of the individual regression lines. While the use of a series of three radiographs to obtain steady state wear (regression lines) has been employed in an attempt to address this issue of variability within the data and account for bedding-in (permanent deformation of the liner without the release of wear particles) with some benefit (McCalden et al., 2009), once again, the relative gains were weighed against access and cost and the decision to have a larger sample was deemed to be of greater importance.

A final limitation in this work is the cross-sectional design of this study which does not permit investigators to make any temporal claims or establish causal links between exposures and outcomes.
Chapter 7 Future Directions

Understanding the temporal relationship between the evolution of clinical parameters (pain, functional or performance limitations) and radiographic markers of potential THR failure will allow researchers and clinicians to develop and target intervention protocols where most effective and cost-efficient and lead to the development of a feasible alternative method of post-operative surveillance.

Larger, longitudinal cohort studies are needed where patients would be followed, starting at 10 to 15 years after THR, for 3 to 5 years. Such studies would require a minimum of three follow-up visits including an x-ray, in order to provide sufficient material to estimate annual linear wear using regression lines (estimating steady state penetration) which is necessary to reduce variability in the Martell wear measurements. While using steady-state wear was impractical within the present context, applying this technique in longer-term follow-up studies might reduce some of the variation created through initial bedding-in, a non-particle releasing permanent deformation of the polyethylene liner, and thus might improve the statistical power of the study.

Evaluation of self-reported pain, self-reported function and walking performance should include the newly identified pain measures (pain after six minute walk test and the ICOAP), administered consistently by interviewer. Standardized x-ray protocols should be developed and employed in a consistent manner by all sites to minimize variability in the outcome measures. Investigators should consider targeted sampling approaches in order to get patients that represent mild, moderate and severe levels of pain or functional limitation to establish an initial link. Since pain and other exposures have been shown to evolve and become more intrusive over time (G.A. Hawker et al., 2008), even though constant pain was not found to be a significant factor in the early follow-up period in the current investigation, it should be closely evaluated after longer post-operative follow-up periods. Provided that there is sufficient statistical power, future investigators should consider controlling for, or stratifying the sample based on implant or polyethylene liner type.

If the present findings and indications for a significant association with six minute walk test pain and the intermittent ICOAP pain are established in future work, it will be critical to
examine whether the relationship is maintained when the influence of pain catastrophizing and other covariates are considered. If the relationship persists when adjusted for this construct and other relevant covariates, further multi-center studies will be needed to confirm the utility of the identified screening tool.

Once a tool is identified, an evaluation and comparison of the diagnostic properties of the candidate screening tools would be required. On the basis of that comparison, the most discriminative candidate tool with high sensitivity (> 0.8) and acceptable specificity (> 0.4) could be evaluated in a prospective longitudinal study to evaluate the predictive properties of the pain scales with failure. This is essential in order to achieve the intended purpose since, in using the measure as a screening tool for post-operative monitoring, current pain would be used as a predictor of a patient’s ‘state-of-risk’. Johnson et al. have described a similar methodology in a 2007 study that looked at the association between changes in WOMAC scores with radiographic indications in a pre-surgical OA population (Johnson et al., 2007).
Chapter 8 Conclusion

In a small sample of primary THR recipients, 4-8 years post surgery for osteoarthritis, few individuals had any evidence of significant self-reported pain or self-reported functional limitation, significant polyethylene wear or significant periprosthetic boney change. These facts severely limited the statistical power of the study to evaluate relationships between these variables. While no statistically significant relationship was found between self-reported pain or function, or measures of walking performance to radiographic markers of potential THR failure (polyethylene wear and periprosthetic boney change), we cannot conclude with confidence that an association does not exist. Trends indicated that there were more individuals with higher annual polyethylene wear rates who also experienced some pain after the six minute walk test or more than mild intermittent pain as measured by the Measure for Intermittent and Constant Osteoarthritis Pain (ICOAP). This is the first study to identify the potential importance of pain after a performance measure (six minute walk test), and the Measure for Intermittent and Constant Osteoarthritis Pain (ICOAP), for future investigations into alternative measures of post-operative surveillance for total hip replacement.
References


*Canadian joint replacement registry (CJRR) 2007 Annual Report—Hip and Knee Replacements in Canada (Ottawa: CIHI, 2008).* Canadian Institute for Health Information.


Appendix 1: Informed Consent
**INFORMED CONSENT TO PARTICIPATE IN A RESEARCH STUDY**

**Study Title:** Safe Activities Following Elective Total Hip Replacement – The “SafeT” Study

**Principal Investigator:** Dr. Hans Kreder, Orthopaedic Surgeon, Sunnybrook Health Sciences Centre

**Funding:** This project is funded by a grant from the Canadian Institutes of Health Research (CIHR). This is the Government of Canada’s health research funding agency.

**BACKGROUND AND PURPOSE OF THIS RESEARCH**

You are being asked to consider participating in a research study. This form explains the purpose of this study, provides information about the study and the tests and procedures involved, possible risks and benefits, and the rights of participants.

The purpose of this research is to study the types of exercises and activities that people participate in who have had a total hip replacement. We are also measuring how often and how hard these patients exercise, and how much pain, if any, they experience when exercising or doing daily activities. More information is needed about which activities are safe and which ones place excess wear on the hip joint. Results of this study will help us develop guidelines for safe activity following total hip replacement.

Please read this form carefully and ask any questions you may have. You may take as much time as you wish to decide whether or not to participate.

**PROCEDURES**

You are being asked to participate in this study because you had a diagnosis of osteoarthritis and have had hip replacement surgery.

We are asking patients who had their hip replacement surgery 5-7 years ago to participate in this study.
If you agree to take part in this study you will come in to see your surgeon twice over a 5-year period – once at the beginning of the study and once at the end.

At the first appointment with your surgeon, you will:
- be asked questions about your current medical state and your medications,
- be asked to complete a few surveys that include questions about your physical activities, pain, function, and mobility,
- have the distance you can walk in six minutes measured.

Then once a year for the next 3 years, you will only need to complete the questionnaires by phone with one of our research assistants. At the end of the 5-year study, we will ask you to come back to see your surgeon, complete the questionnaires, and do the six-minute walk test again.

We will also review your medical records to confirm the information that you give us and decrease the amount of time you spend with the researchers answering questions.

**HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?**
It is anticipated that 1,200 people will participate in this study at Sunnybrook Health Sciences Centre’s two locations, including the Holland Orthopaedic & Arthritic Centre. The length of this study for participants is 5 years.

**BENEFITS**
We do not know if you personally will benefit from participating in the study. The results of the study may add new information that may help identify safe activities for hip replacement patients. This information will benefit future hip replacement patients by providing them with safer activity guidelines for their recovery and life after hip surgery.

**RISKS/DISCOMFORT/POTENTIAL HARMs**
There are no risks to participating in this study.

**ALTERNATIVES TO PARTICIPATING**
You will receive usual medical care by your doctor if you choose not to participate in this study.
CAN PARTICIPATION IN THIS STUDY END EARLY?
The investigator may decide to remove you from this study without your consent if you are unable or unwilling to follow the study procedures. You can also choose to end your participation at anytime. If you withdraw voluntarily from the study you are encouraged to contact Monica Kunz, Research Coordinator, Orthopaedic Surgery, Sunnybrook Health Sciences Centre at (416) 967-8616 immediately.

WHAT ARE THE COSTS OF PARTICIPATING IN THIS STUDY?
Although you will not have to pay for additional visits to your surgeon, you will have to pay for any transportation, parking or transit costs you incur. You will not be paid to participate in this study.

CONFIDENTIALITY AND RIGHTS OF RESEARCH SUBJECTS
When you consent to participate in this study, you will be assigned a subject number for identification purposes. At the end of the study the data will be analyzed and the findings published in a medical journal. Your name and participation will always be kept confidential and will never appear in any study databases or medical journals.

RIGHTS OF SUBJECTS
You are taking part in this study of your own free will. You can refuse to take part right now or can stop taking part at any time. Your decision to take part or not take part in this study will have no effect on the quality of your medical care. Your legal rights are not affected by participating in the study.

FURTHER QUESTIONS
You will be given a copy of this consent form and if you have any questions about the research before or after participation, you may contact the Principal Investigator, Dr. Hans Kreder at Sunnybrook Health Sciences Centre at (416) 480-6816, or Study Coordinator, Monica Kunz at (416) 967-8616. If you have questions about your rights as a research participant or any ethical issues related to this study that you wish to discuss with someone not directly involved in the study, you may call Dr. Philip C. Hebert, Chair of the Sunnybrook Research Ethics Board, at (416) 480-4276
DOCUMENTATION OF INFORMED CONSENT
Study Title: Safe Activities Following Elective Total Hip Replacement – The “SafeT” Study

Participant:
By signing this form, I confirm that:
• This research study has been fully explained to me, and all of my questions answered to my satisfaction.
• I understand the requirements of participating in this research study.
• I realize that I will be completing questionnaires and a 6-minute walk test in addition to my usual care.
• I have been informed of the risks and benefits, if any, of participating in this study.
• I have been informed of any alternatives to participating in this research study.
• I have been informed of the rights of research participants.
• I know that I may ask now, or in the future, any questions I have about the study or the research procedures.
• I have read each page of this form.
• I authorize access to my personal health information (medical record) and research study data as explained in this form.
• I have been assured that records relating to my care will be kept confidential and that no information will be released or printed that would disclose my personal identity without my permission.
• I understand my participation is completely voluntary and that I am free to withdraw from the study at any time. I further understand that if the study is not completed, or if I withdraw from it at any time, the quality of my medical care at Sunnybrook Health Sciences Centre will not be affected.
• I have agreed to participate in this study.

___________________________________________________
Name of participant (print)

______________________________________________  _____________________
Signature Date

Person obtaining consent:
By signing this form, I confirm that:
• I have explained this study and its purpose to the participant named above.
• I have answered all questions asked by the participant.
• I will give a copy of this signed and dated document to the participant.

__________________________________________________
Name of person obtaining consent (print)

______________________________________________  _____________________
Signature Date
Appendix 2: Demographics Form
SH-02-7

Study ID number: Patient’s initials (F-M-L) Date of Baseline Visit (month, day, year, i.e. January 1, 2008) Years since surgery

Patient Eligibility Inclusion Criteria: If YES to all, then patient can be included.

Was the patient be ≤ 80 years old at time of surgery? ☐ Yes ☐ No
Does the patient speak English? ☐ Yes ☐ No
Is the patient capable of providing informed consent and able to understand the survey questions? ☐ Yes ☐ No
Does the patient have osteoarthritis? ☐ Yes ☐ No
(Patients with congenital hip dysplasia with a secondary diagnosis of O/A at time of surgery are eligible. Patients with a history of gout or avascular necrosis can be included.)

Patient Eligibility Exclusion Criteria: If YES to one or more at the time of surgery, then patient is excluded.

Trauma ☐ Yes ☐ No
Previous hip fracture (in hip being studied) ☐ Yes ☐ No
Metastatic cancer ☐ Yes ☐ No
Significantly impaired function and mobility due to severe inflammatory arthritis or systemic disease requiring disease modifying medications (i.e. severe RA, ankylosing spondylitis, polymyalgia rheumatica, parkinson’s) ☐ Yes ☐ No
Rheumatoid Arthritis requiring medication ☐ Yes ☐ No
Paget’s Disease or Lupus ☐ Yes ☐ No
(Patient is eligible if any of these conditions was diagnosed after surgery took place.)

Demographic Information

☐ Female ☐ Male

Current: Pre-op:
Height: ________ cm Height: ________ cm
Weight: ________ kg Weight: ________ kg

Birthdate: ________/______/_________ Age now: ________ Age at surgery: ________
MONTH DAY YEAR

Current Marital Status:
☐ Married/common-law ☐ Single ☐ Widowed ☐ Divorced

Current usual occupation: (please check all that apply):
☐ Employed ☐ Retired ☐ Unemployed ☐ Disabled ☐ Student
☐ professional, technical, managerial ☐ clerical, admin
☐ service sector, sales ☐ homemaker
☐ agriculture, fishery, forestry and related ☐ factory
☐ trades (machine, plumbing, carpentry, etc.)
☐ other ______________________________

Pre-retirement occupation? _______________________________________

Recorded by: ______________________________
Subject I.D. # SH-02-7
Comorbidity Questions:

patient should answer both columns →

| A | Rheumatoid arthritis (swollen joints) | Yes | No | Yes | No |
| B | Osteoarthritis otherwise described as degenerative arthritis, or wear and tear arthritis (the common kind with aging) | Yes | No | Yes | No |
| C | Other types of arthritis such as lupus, scleroderma, gout, fibromyalgia, psoriatic | Yes | No | Yes | No |
| D | Persistent back or neck problems | Yes | No | Yes | No |
| E | Lung problems (emphysema, chronic bronchitis, asthma) | Yes | No | Yes | No |
| F | High blood pressure | Yes | No | Yes | No |
| G | Heart problems (angina, heart attack, heart failure) | Yes | No | Yes | No |
| H | Hardening of the arteries in the legs (arteriosclerosis) | Yes | No | Yes | No |
| I | Amputation of a leg | Yes | No | Yes | No |
| J | Stomach ulcer – IF YES | Yes | No | Yes | No |
| 1. | Has this condition been diagnosed by endoscopy (where a doctor looks into your stomach through a scope) or an upper GI or barium swallow (where you swallow chalky dye and then x-rays are taken)? | Yes | No | Yes | No |
| 2. | Has a physician ever told you that this stomach ulcer was considered to be a bleeding ulcer? | Yes | No | Yes | No |
| 3. | Were you ever hospitalized for this condition? | Yes | No | Yes | No |
| 4. | Are you taking any medication for this problem? | Yes | No | Yes | No |
| K | Kidney disease (kidney failure) | Yes | No | Yes | No |
| L | Cancer (not skin cancer) | Yes | No | Yes | No |
| M | Major paralysis or neurologic problems, such as stroke, MS, Parkinson’s | Yes | No | Yes | No |
| N | Diabetes | Yes | No | Yes | No |
| O | Depression | Yes | No | Yes | No |
| P | Other major mental illness (other than depression) | Yes | No | Yes | No |
| Q | Osteoporosis (brittle bones) | Yes | No | Yes | No |
| R | Cirrhosis of the liver | Yes | No | Yes | No |

History of major injuries/accidents as an adult (i.e. car accident, workplace, other major surgery, major sports injury):

Mobility Aids: □ none □ 1 cane □ 2 canes □ walker □ other: __________

Revision surgery information:

Is revision surgery planned or already scheduled? □ Yes □ No

If Yes, date of future revision surgery:

Reason for revision:

□ Aseptic loosening □ Poly wear □ Osteolysis
□ Instability □ Malposition □ Implant fracture
□ Bone fracture □ Infection □ Reattach trochanter
□ Other: ____________________________

SafeT SH cohort 2 Baseline data form.doc
Rev. Apr 6 09 Page 2 of 3

Recorded by: ____________________________
Subject I.D. # SH-02-7
Patient’s history of **high-level** physical activity (get from patient): (e.g. competitive sports and teams, long distance running, master’s swim club, regular league participation, etc.)

Please check all that apply and provide details:  ☐ none to report

- ☐ Childhood sports team/activity (specify): ________________________________
- ☐ High school team/activity (specify): ________________________________
- ☐ University/college sports team/activity (specify): ________________________________
- ☐ Adult sports/activity (specify): ________________________________

**Physical Function Data**

**Surveys:**
- ☐ Physical Activity (blue) survey (12 months prior to today)
- ☐ Survey booklet (blue) completed today in clinic
- ☐ Surveys scheduled to complete over the phone on (date): _______________

**6-Minute Walk Test:** (stop watch and clipboard needed)
Holland Centre track is 46 metres long, following the yellow or blue line. Whatever block they stop in, count that block as one whole metre (i.e. do not use decimal points for partial metres walked).
At Sunnybrook, use the portable, green Rolatape wheel and round up to the nearest metre. One lap around the Orthopaedic Surgery clinic at SB is 50.5 metres. Use the Rolatape for the last partial lap.

Instruct patient to “**Cover as much ground as you can in 6 minutes. Take a rest whenever you need one and then, when you feel ready, continue.**” Offer a little encouragement, but not so much that you distract the patient while he or she is walking. i.e. “**Good job, you have 2 minutes left.**”

If you walk part of the way with the patient, walk behind, not beside or in front of them.

Lap count: ________________ Plus _______ metres

Metres walked: ________________ # rest breaks: ________________

Pain score (circle one): [ ] 0 [ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] 7 [ ] 8 [ ] 9 [ ] 10

☐ 6MW Test – unable to complete

6MW Test other comments: ________________________________

**Medications:**

Are you taking anything for your hip pain?
- ☐ Yes  ☐ No  ☐ Don’t know
If yes, list what: ____________________________________________

Do not record multivitamins, calcium, or other vitamins and minerals.
Appendix 3: Surgery Data Form
### The SafeT Study (Safe Activities Following Elective Total Hip Replacement)
#### Both Cohorts: SURGICAL DATA FORM

<table>
<thead>
<tr>
<th>SH-0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study ID number</strong></td>
</tr>
</tbody>
</table>

2. **Surgeon:**
- Cameron, H
- Cameron, J
- Gollish
- Holtby
- Jenkinson
- Wright
- Kreder
- Mumaghan
- Wadey
- Nousiainen
- Other

3. **Side:**
- Left
- Right

6. **American Society of Anesthesiologists Physical Status Classification System:**
- P1 – normal healthy
- P2 – mild systemic disease
- P3 – severe systemic disease
- P4 – severe systemic disease that is a constant threat to life
- N/A

8. **Primary diagnosis:**
- O/A
- Other

9. **Secondary diagnosis:**
- None
- O/A
- Osteonecrosis
- Inflammatory arthritis
- Post-traumatic osteoarthritis
- Acute hip fracture
- Childhood hip problem
- Old hip fracture
- Tumour
- Infection
- Other

13. **Approach:**
- Smith/Peterson
- Ant-lat
- Direct-lat
- Post-lat
- MIS – ant
- MIS - 2 incision

14. **Acetabular Liner:**
- Cemented
- Uncemented

15. **Acetabular Liner:**
- Polyethylene
- Metal
- Ceramic
- Other

17. **Femoral Head:**
- Polyethylene
- Metal
- Ceramic
- Other

19. **Main Femoral Stem:**
- Cemented
- Uncemented

20. **Circlage wires:**
- Used
- Not used (if no stickers with other hard wear then no)

**Surgical complications:** (see chart In-patient notes—OR Notes, Recovery Room Notes, In-patient Progress Notes; and Out-patient Record Sheet): No complications

- DVT (deep vein thromb.)
- Nerve injury
- Heterotopic ossification
- UTI
- Allergic to implant
- Infection
- Myocardial infarction
- Fat embolism
- Pulmonary embolus
- Arterial injury
- Respiratory depression
- Renal failure
- Implant dislocation
- Leg length discrepancy
- Anesthetic complications
- Pneumonia

Other:

**Discharge date:**
- Month
- Day
- Year
- L.O.S. days

- to FIT Program
- Home
- Rehab
- Nursing Home
- Retirement Home

Recorded by: _______________________

SafeT SH Both Cohorts Surgery Data Form.doc Rev. Nov 25 09

---

The SafeT Study (Safe Activities Following Elective Total Hip Replacement)
Appendix 4: Questionnaires
Safe Activities Following Elective Total hip replacement

Thank you for agreeing to participate in this research study. Please answer all of the questions in this booklet. Instructions for specific sections are provided.

All of the information you provide will be held in strict confidence. No data will be reported that will identify you, or link an individual to specific or general responses.

The SafeT research study is being conducted at Sunnybrook Health Sciences Centre (both the Sunnybrook and Holland Orthopaedic & Arthritic campuses), and London Health Sciences Centre. It is funded by the Canadian Institutes of Health Research.
OSTEOPATHRITIS Pain ASSESSMENT TOOL – HIP Joint

People have told us that they experience different kinds of pain (including aching or discomfort) in their hip. To get a better sense of the different types of hip pain you may experience, we would like to ask you about any “constant pain” (pain you have all the time) separately from any pain that you may experience less often, that is, “pain that comes and goes”. The following questions will ask you about the worst/most troublesome pain that you have experienced in your hip in the PAST WEEK.

A) CONSTANT PAIN (pain that is there all the time but may vary in intensity)

For each of the following questions, please select the response that best describes, on average, your constant hip pain in the PAST WEEK.

1. In the past week, how intense has your constant hip pain been?
   - Not at all/ No constant hip pain
   - Mildly
   - Moderately
   - Severely
   - Extremely

2. In the past week, how much has your constant hip pain affected your sleep?
   - Not at all/ No constant hip pain
   - Mildly
   - Moderately
   - Severely
   - Extremely

3. In the past week, how much has your constant hip pain affected your overall quality of life?
   - Not at all/ No constant hip pain
   - Mildly
   - Moderately
   - Severely
   - Extremely

4. In the past week, how frustrated or annoyed have you been by your constant hip pain?
   - Not at all/ No constant hip pain
   - Mildly
   - Moderately
   - Severely
   - Extremely

5. In the past week, how upset or worried have you been by your constant hip pain?
   - Not at all/ No constant hip pain
   - Mildly
   - Moderately
   - Severely
   - Extremely

Page 2 of 7
B) PAIN THAT COMES AND GOES (pain that is not there all the time)

For each of the following questions, please select the response that best describes your *hip pain that comes and goes*, on average, in the PAST WEEK.

6. In the past week, how intense has your most severe *hip pain that comes and goes* been?

- □, Not at all/
- □, No hip pain that comes and goes
- □, Mildly
- □, Moderately
- □, Severely
- □, Extremely

7. In the past week, how frequently has this *hip pain that comes and goes* occurred?

- □, Never /
- □, No hip pain that comes and goes
- □, Rarely
- □, Sometimes
- □, Often
- □, Very often

8. In the past week, how much has your *hip pain that comes and goes* affected your sleep?

- □, Not at all/
- □, No hip pain that comes and goes
- □, Mildly
- □, Moderately
- □, Severely
- □, Extremely

9. In the past week, how much has your *hip pain that comes and goes* affected your overall quality of life?

- □, Not at all/
- □, No hip pain that comes and goes
- □, Mildly
- □, Moderately
- □, Severely
- □, Extremely

10. In the past week, how frustrated or annoyed have you been by your *hip pain that comes and goes*?

- □, Not at all/
- □, No hip pain that comes and goes
- □, Mildly
- □, Moderately
- □, Severely
- □, Extremely

11. In the past week, how upset or worried have you been by your *hip pain that comes and goes*?

- □, Not at all/
- □, No hip pain that comes and goes
- □, Mildly
- □, Moderately
- □, Severely
- □, Extremely
### Pain Catastrophizing Scale

Please think about a previous painful experience and indicate to which degree you recall experiencing the following thoughts and feelings during the episode of pain (please circle the appropriate number).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Not at all</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) I worry all the time about whether the pain will end.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2) I feel I can’t go on.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3) It’s terrible and I think it’s never going to get any better.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4) It’s awful and I feel that it overwhelms me.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5) I feel I can’t stand it anymore.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6) I become afraid that the pain may get worse.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7) I think of other painful experiences.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8) I anxiously want the pain to go away.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9) I can’t seem to keep it out of my mind.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10) I keep thinking about how much it hurts.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>11) I keep thinking about how badly I want the pain to stop.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>12) There is nothing I can do to reduce the intensity of the pain.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>13) I wonder whether something serious may happen.</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
### NRS for Arthritis Pain

How would you rate the average intensity of your arthritis pain during the PAST WEEK on a 0-10 scale, where 0 is ‘no pain’ and 10 is ‘worst pain possible.’

Circle only ONE number.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Pain</td>
<td>Worst Possible Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Pain</td>
<td>Worst Possible Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Pain</td>
<td>Worst Possible Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
WOMAC™ Osteoarthritis Index

Section A – PAIN

Instructions to patients:
The following questions concern the amount of PAIN you are currently experiencing due to arthritis in your hip. For each situation please enter the amount of pain experienced in the last 48 hours. (Please mark your answers with an “X”)

QUESTION: How much pain do you have?

1. Walking on a flat surface
   - None
   - Mild
   - Moderate
   - Severe
   - Extreme

2. Going up or down stairs
   - None
   - Mild
   - Moderate
   - Severe
   - Extreme

3. At night while in bed
   - None
   - Mild
   - Moderate
   - Severe
   - Extreme

4. Sitting or lying
   - None
   - Mild
   - Moderate
   - Severe
   - Extreme

5. Standing upright
   - None
   - Mild
   - Moderate
   - Severe
   - Extreme

Section B – STIFFNESS

Instructions to patients:
The following questions concern the amount of joint STIFFNESS (not pain) you have experienced in the last 48 hours in your hip. Stiffness is a sensation of restriction or slowness in the ease with which you move your joints. (Please mark your answers with an “X”)

How severe is your stiffness when you wake up in the morning?

- None
- Mild
- Moderate
- Severe
- Extreme

How severe is your stiffness after sitting, lying or resting later in the day?

- None
- Mild
- Moderate
- Severe
- Extreme
**Section C – FUNCTION**

**Instructions to patients:**

The following questions concern your **PHYSICAL FUNCTION**. By this we mean your ability to move around and to look after yourself. For each of the following activities, please indicate the degree of difficulty you have experienced in the **last 48 hours** due to arthritis in your hip.

(Please mark your answers with an X)

**QUESTION: What degree of difficulty do you have with:**

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
<th>Extreme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Descending (going down) stairs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Ascending (going up) stairs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Rising from sitting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Standing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Bending to floor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Walking on a flat surface</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Getting in/out of a car or getting on/off a bus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Going shopping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Putting on socks/stockings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Rising from bed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Taking off socks/stockings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Lying in bed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Getting in/out of the bath</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Sitting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Getting on/off the toilet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Heavy domestic duties (such as mowing the lawn, lifting heavy grocery bags, vacuuming)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Light domestic duties (such as tidying a room, dusting, cooking)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**THANK YOU FOR COMPLETING THESE QUESTIONS**