EFFECT OF DELAYS ON HEALTH RELATED QUALITY OF LIFE OUTCOME AFTER PRIMARY TOTAL HIP ARTHROPLASTY

By

MIN XU

M.B., Tongji Medical University, 1999

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

MASTER OF SCIENCE

In

THE FACULTY OF GRADUATE STUDIES

Department of Health Care and Epidemiology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 2004

© Min Xu, 2004
Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Min Xu
Name of Author (please print)

[Signature]
Date (dd/mm/yyyy)

Title of Thesis: Effect of delays on health related quality of life after primary total hip arthroplasty

Degree: M.Sc Year: 2004

Department of Health care and epidemiology
The University of British Columbia
Vancouver, BC Canada
Abstract

Objective
The main purpose of this research is to investigate the relationship between the waiting time for elective total hip arthroplasty (THA) and changes in the health-related quality of life (HRQOL) outcomes measured at the time of registration on the wait list and one year after the operation. The primary research objective is to assess whether expedited access to THA results in a larger proportion of patients showing “better than expected” lower-extremity function 12 months after the operation.

Methods
This prospective cohort study was conducted at the Vancouver Hospital & Health Sciences Center. Patients who entered the waiting list for primary THA with osteoarthritis were included in the study. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire was used to assess each patient at the surgical consultation time (baseline) and one year after the operation (follow up). Log-linear regression models were used to determine whether an individual patient achieved a “better than expected” outcome. Multivariate regression models were used to assess the relationship between waiting time and the probability of achieving a “better than expected” outcome.

Results
We studied 147 patients at the baseline and at the follow-up assessments. The average waiting time was 6.3 months with a standard deviation of 4.4 months. We found that the baseline WOMAC score is a very significant predictor for the follow up WOMAC score in function (p=0.0005), pain (p=0.0036), and stiffness (p=0.0004). Using an individualized expected outcome based on baseline score, we determined whether a patient experienced the expected achievement through surgery. We found that waiting time was significantly associated with the probability of achieving a “better than expected” function outcome (p=0.05).

Conclusions
Expeditied access to THA results in a larger proportion of patients showing “better than expected” lower-extremity function at 12 months after the operation. Shorter waiting time is associated with the increased probability of achieving a “better than expected” outcome for an individual patient.
TABLE OF CONTENTS

Abstract ........................................................................................................ II
Table of Contents ......................................................................................... III
List of Tables ................................................................................................ IV
List of Figures ............................................................................................... V
Acknowledgements ...................................................................................... VI

1 INTRODUCTION ................................................................................. 1
  1.1 Purpose .......................................................................................... 1
  1.2 Hypothesis ..................................................................................... 2
  1.3 Objectives ...................................................................................... 2
  1.4 The following chapters .................................................................... 3

2 BACKGROUND .................................................................................. 4
  2.1 Osteoarthritis (OA) ...................................................................... 4
  2.2 Total hip arthroplasty (THA) ......................................................... 5
  2.3 Health related quality of life (HRQOL) ....................................... 8
  2.4 Responsiveness ............................................................................. 9
  2.5 Waiting time ................................................................................ 16

3 METHODS .......................................................................................... 22
  3.1 Overview ..................................................................................... 22
  3.2 Study design ................................................................................ 22
  3.3 Statistical analysis ....................................................................... 27

4 RESULTS ............................................................................................ 35
  4.1 Study population ......................................................................... 35
  4.2 Waiting time ............................................................................... 36
  4.3 Patient characteristics ................................................................. 37
  4.4 Non-responder and responder comparison .................................. 39
  4.5 WOMAC function ....................................................................... 41
  4.6 WOMAC pain .............................................................................. 54
  4.7 WOMAC stiffness ...................................................................... 65

5 DISCUSSION AND CONCLUSION .................................................. 75
  5.1 Overview ..................................................................................... 75
  5.2 Discussion ................................................................................... 75
  5.3 Limitations .................................................................................. 79
  5.4 Conclusions ............................................................................... 80
  5.5 Future studies ............................................................................. 81

BIBLIOGRAPHY ..................................................................................... 82
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Summary table of literature</td>
<td>21</td>
</tr>
<tr>
<td>Table 2</td>
<td>Patient demographic statistics</td>
<td>38</td>
</tr>
<tr>
<td>Table 3</td>
<td>Comparison of responders and non-responders</td>
<td>40</td>
</tr>
<tr>
<td>Table 4</td>
<td>Parameter estimation for the log-linear regression model</td>
<td>46</td>
</tr>
<tr>
<td>Table 5</td>
<td>Relationship between waiting time and one-year outcome measured by WOMAC function scores</td>
<td>51</td>
</tr>
<tr>
<td>Table 6</td>
<td>Parameter estimation for the log-linear regression</td>
<td>59</td>
</tr>
<tr>
<td>Table 7</td>
<td>Relationship between waiting time and one-year outcome measured by WOMAC pain scores</td>
<td>63</td>
</tr>
<tr>
<td>Table 8</td>
<td>Parameter estimation for the log-linear regression model</td>
<td>69</td>
</tr>
<tr>
<td>Table 9</td>
<td>Relationship between waiting time and one-year outcome measured by WOMAC stiffness scores</td>
<td>73</td>
</tr>
</tbody>
</table>
List of Figures:

Figure 1: Schema of patient flow from referral to surgery........................................... 24
Figure 2: The distribution of waiting time (in months)................................................. 36
Figure 3: The distribution of baseline WOMAC function scores................................. 41
Figure 4: The distribution of follow-up WOMAC function scores............................... 42
Figure 5: Baseline WOMAC function scores and follow-up WOMAC function scores.................................................. 44
Figure 6: The distribution of the logarithms of WOMAC function scores post THA... 45
Figure 7: The regression line for the logarithms of follow-up scores versus baseline WOMAC function scores................................................. 48
Figure 8: 90% confidence interval for the median of expected function outcome..... 49
Figure 9: The percent achieving a “better than expected” functional outcome in short and long waiting groups................................................................. 50
Figure 10: Residuals for the logistic regression model................................................. 52
Figure 11: The distribution of baseline WOMAC pain scores...................................... 54
Figure 12: The distribution of follow-up WOMAC pain scores................................... 55
Figure 13: Baseline WOMAC pain scores and follow-up WOMAC pain scores........ 57
Figure 14: The distribution of the logarithms of follow-up WOMAC pain scores........ 58
Figure 15: The percent achieving a “better than expected” pain outcome in short and long waiting groups................................................................. 62
Figure 16: The distribution of baseline WOMAC stiffness scores.............................. 65
Figure 17: The distribution of follow-up WOMAC stiffness scores............................ 66
Figure 18: Logarithms of follow up WOMAC stiffness scores.................................... 68
Figure 19: The percent achieving a “better than expected” stiffness outcome in short and long waiting groups................................................................. 72
Acknowledgements

Many people have helped me during the development of this thesis. First of all, I would like to thank my major professor Dr. Boris Sobolev, who provided a motivating, enthusiastic atmosphere during our many discussions. It was a great pleasure to conduct this thesis under his supervision. I also want to thank Dr. Donald Garbuz, my second supervisor, for his guidance and encouragement. I would especially like to thank my supervisory committee member, Dr. Jacek Kopec, who provided constructive comments during my thesis time. I would also like to thank Dr. Matthew Liang whose suggestions were substantial for this research and Dr. Stephen Marion for his encouragement during my research.

Needless to say, I am very grateful to my colleagues in the Arthritis Research Center of Canada for their support (and patience). Special thanks to James Latteier for reading draft versions of the thesis and his comments during our discussions. Many thanks to Francisco Luna for his valuable suggestions.

I would also like to thank my mother and sister for their support.
1 INTRODUCTION

1.1 Purpose

Primary total hip arthroplasty (THA) is an effective treatment for severe osteoarthritis (OA) of the hip. The operation is primarily done to relieve pain and improve health-related quality of life (HRQOL). When treatment is delayed, the condition of a patient on a surgical wait list may deteriorate. Treatment delay may also compromise long-term outcomes. However, the question of whether prompt access to surgery can benefit patients over the long term has attracted little research.

The main purpose of this research is to investigate the relationship between the duration of wait for elective THA and changes in HRQOL outcomes measured at the time of registration on the wait list and one year after the operation. The evidence generated will be used to address the issue of safe waiting periods.

To describe within-individual change over time, this research focuses on the individual patterns of change in HRQOL outcomes before and after THA. Using a regression method, we estimate an expected postoperative HRQOL score for a certain individual preoperative score. To evaluate whether waiting time for surgery predicts changes in HRQOL outcome, we study inter-individual change. For this we classify individual postoperative outcome as “better than expected” or “not better than expected” based on the postoperative score expected for a given baseline score.
1.2 Hypothesis

It is biologically plausible that prolonging the arthritic process in the hip may result in muscle atrophy, tissue contractures, and deterioration of health status that may not be fully recoverable post surgery. Therefore, we hypothesize that longer waiting time for surgery may be negatively associated with postoperative outcomes.

- Primary hypothesis: that prolonged waiting is a significant independent risk factor for achieving a “better than expected” functional outcome following THA.
- Secondary hypothesis: that prolonged waiting is a significant independent risk factor for achieving a “better than expected” pain and stiffness outcome following THA.

*Functional outcome:* patients’ ability to move about and look after themselves

*Pain outcome:* pain that patients experience in the hip

*Stiffness outcome:* the sensation of restriction or slowness in the joints

1.3 Objectives

The primary research objective is to assess whether expedited access to THA results in a larger proportion of patients showing a “better than expected” HRQOL score for lower-extremity function 12 months after the operation. Related objectives are to assess whether expedited access to THA results in a larger proportion of patients showing a “better than expected” HRQOL score for pain and lower-extremity stiffness 12 months after the operation. Data in response to this question will enable the assessment of long-term treatment outcomes in relation to the duration of wait for surgery. In the long run, we aim to develop an algorithm to determine optimal access time based on the severity of OA.
1.4 The following chapters

Chapter 2 reviews the literature and provides evidence from previous studies on the biological plausibility of the effect of delay on HRQOL outcomes after primary THA. Chapter 2 also shows that controversial conclusions may be drawn from previous studies. The gap in the literature prompted the development of the present study. Chapter 3 describes the details of the study design, the study process, the method for data collection and data analysis. Chapter 4 focuses on the results of the data analysis. Chapter 5 draws conclusions and discusses the results. Chapter 5 also comments on the limitations of the study and suggests directions for future research.
2 BACKGROUND

2.1 Osteoarthritis (OA)

OA is a chronic, progressive form of arthritis. It usually affects the large weight-bearing joints of the hips and knees, but it is also found in the hands, feet, and spine. The water content of the cartilage increases in OA joints while the protein makeup of cartilage degenerates. In advanced cases, the cartilage between the bones of the joint can be totally lost. OA patients suffer from pain and loss of function as the disease progresses. It is a major public health problem in Canada and around the world. A World Health Organization (WHO) report ranked OA as one of the ten most important causes of disability-adjusted life years lost, ahead of such conditions as diabetes and breast cancer.\(^1\)

The Arthritis Society estimates that approximately 10% of Canadians suffer from OA.\(^2\)

The symptoms of hip osteoarthritis usually develop very slowly, though in a few cases pain starts abruptly. Pathological lesions can precede symptoms by years, and the pain may start with minor trauma to an already diseased joint. The early diagnosis can be difficult, for the pain can vary greatly in both its site and nature. It may be felt in the area of the buttock, groin, thigh, or knee, and its character may vary from a dull ache to sharp stabbing pains. In the early stage of OA, the pain is generally associated with activity, and exercise may induce bouts of pain that last for several hours. As the disease progresses, severe pain may be present at night or during rest. Stiffness of the hip is usual,
particularly after inactivity, and can be the presenting feature—for example, patients may complain of difficulty in putting on socks and shoes.\textsuperscript{3}

For most patients, the long-term outcomes are poor, though many cases remain relatively stable for years. Patients who require surgery for hip osteoarthritis often complain of disease deterioration and severe symptoms over a relatively short period (1-2 years). Some features associated with a relatively good or poor outcome have been identified: elderly patients and those with severe pain or discrepancies in leg length are more likely to show rapid progression of the disease, as well as those with radiographic evidence of a superolateral subluxation of the hip or a paucity of bone response.\textsuperscript{4} Up to 5% of advanced cases of hip osteoarthritis improve spontaneously with radiographic evidence of improvement as well as lessening pain, revealing the potential for repair in patients with concentric pattern hip osteoarthritis and extensive radiographic changes.\textsuperscript{5}

\subsection{2.2 Total hip arthroplasty (THA)}

\subsubsection{2.2.1 Total hip arthroplasty}

THA is the most effective treatment for patients with severe OA in terms of improved physical function and pain relief.\textsuperscript{5,6,7,8,9,10} Bachmeier et al. reported that WOMAC measures improved significantly one year after THA: physical function (68%), pain (71%), and stiffness (55%).\textsuperscript{11} This study also found that SF-36 measures in those having hip surgery improved significantly: 222\% reduction in pain, 247\% improvement in physical function, and 402\% improvement in physical role functioning.
But complications do occur such as infection, dislocations, and long-term biomaterial failure. The mortality rate within 90 days after primary THA is 1.0%. The rate of pulmonary embolus after primary THA is 0.9%. The rate of wound infection after primary THA is 0.2%. The rate of hip dislocation is 3.1% after primary THA and the hospital readmission rate after primary THA is 4.6%. Revision total hip replacement has a higher rate of complications. The mortality rate after revision THA was 2.6%. The rate of pulmonary embolus after revision THA was 0.8%. The rate of wound infection after revision THA was 0.95%. The rate of hip dislocation was 8.4% after revision THA and the hospital readmission rate after revision THA was 10.0%. Factors associated with an increased risk of an adverse outcome include age (the elderly are at a higher risk), gender (men are at a higher risk than women), race (blacks at higher risk than whites), medical co-morbidity, and lower income.

Liang et al. reported a large variation in treatment effectiveness. The current understanding of this variation is limited mostly to patient- and implant-related factors. The role of variations in service delivery practices remains unclear.

2.2.2 Patient related factors

2.2.2.1 Age

Nilsdotter et al. reported that age is a predictor of outcome after total hip replacement for osteoarthritis. One hundred twenty-four patients fulfilled the study criteria (age 50 years
or over and unilateral THA for OA during the study period). This study reports that the age difference of patients receiving THA did not determine their preoperative status. However, younger patients regained more function postoperatively than older patients and reached higher absolute mean SF-36 values, except for pain. At least one year is required for the average OA patient to gain the full benefits of THA. Roder et al. studied 48,000 THAs. This study also found that age has a strong effect on functional and pain outcome after THA.\textsuperscript{17} MacWilliam et al. found that age is a statistically significant risk factor for the change between preoperative and postoperative pain outcomes but was not significant for function.\textsuperscript{18} Jones et al. reported a community-based cohort study that included 197 patients who received primary THA.\textsuperscript{19} In this study, age was not found to be a statistically significant risk factor for function or pain outcomes six months after surgery.

\subsection*{2.2.2.2 Gender}

McMurray et al. suggested that women may need a longer time to recover physically.\textsuperscript{20} Jones et al. reported that, regardless of age, gender might be a risk factor for the change between preoperative and postoperative pain outcomes, but not significant for function.\textsuperscript{19}

\subsection*{2.2.2.3 Bilateral disease}

Roder et al. indicated that bilateral disease is a strong risk factor for postoperative functional outcomes.\textsuperscript{17} The study used the Charnley classification to classify patients into
three classes. Improvement in walking ability was also strongly associated with the Charnley classification, which measures co-morbidity.

2.2.2.4 Co-morbidity

Mahomed et al. indicated that patients with medical co-morbidity were at higher risk of death than those without (OR 2.06, 95% CI 1.59-2.66).\textsuperscript{14} Patients with medical co-morbidity and those with lower income had a higher rate of readmission to the hospital. The presence of medical co-morbidity also increased the odds of hip dislocation and infection.

2.3 Health related quality of life (HRQOL)

The goals of THA include relieving symptoms, minimizing disability and handicap, and reducing the risk of disease progression. In previous orthopedic practice, the patient’s perspective received less attention than the objective measures of disease and impairment.\textsuperscript{21,22} Nevertheless, during the past decade, the orthopaedic community has shifted toward the inclusion of patient-based measures in outcome assessments, thereby expanding its understanding of the effects of intervention.\textsuperscript{23} Decision-making studies on OA treatment have focused increasingly on HRQOL as an important outcome variable. Many randomized, placebo-controlled trials for the effectiveness of treatment use HRQOL measures as valid and useful endpoints.
2.3.1 Disease-specific questionnaire

The Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) is recommended for OA-specific outcomes. It contains dimensions for pain (5 items), stiffness (2 items), and function (17 items). Dimensions are equally weighted and reported as sums, the higher number indicating a greater burden of OA. At present it is the most frequently used measure of pain and self-rated disability among arthroplasty patients. It has been proven that the WOMAC is responsive to changes over time, which makes it suitable for subtle discrimination during waiting. The test-retest reliability of the instrument is between 0.73 and 0.96. Minimal perceptible clinical improvement in individual patients was found to be 9.7, 9.3 and 10.0 for the subscales of pain, physical function, and stiffness respectively.

2.4 Responsiveness

2.4.1 Taxonomy for responsiveness

Beaton et al. distinguished the different kinds of change that can be quantified in a study of responsiveness. The types of change they quantified include:

1. “Minimum potentially detectable change by the instrument
2. Minimum change detectable given the measurement error of the instrument
3. Observed change measured by the instrument in a given population
4. Observed change in a population deemed to have improved
5. Observed change in those deemed to have had an important improvement”
2.4.2 Different perspectives on important changes

The determination of whether an important or satisfactory change has occurred can be based upon a variety of perspectives.27 The clinician can use criteria, such as test results, assessment findings, and general impression to determine the degree of improvement. Patients’ ratings may be based on their personal expectations of the outcome, and satisfaction with the outcome. Societal ratings may be based on whether patients have returned to their social or work roles. Different perspectives may result in different values on the cut off points for important change.

2.4.3 Group level measurements

Most orthopedic research on the effectiveness of THA uses distribution-based approaches for HRQOL post-operation outcome. The utilization of distribution-based approaches to determine the significance of change are based on the statistical characteristics of the obtained sample. Paired t-statistics are most commonly used in a one-group repeated measures design to determine the statistical significance of the change.28 The problem with using the t-test as a measure of change is that it focuses exclusively on the statistical significance of the differences instead of the magnitude of the change itself. In addition, the significance increases with sample size.29

Effect size (ES) and standard response mean (SRM) are also common measures for responsiveness at the group level. But these measurements may be influenced by the heterogeneity of the sample. Characteristics of the baseline distribution strongly influence
the effect size, while variability of the change in the sample may influence the standard response mean.

2.4.4 Measuring individual changes based on sample variations

The measurement of individual changes is an increasingly attractive method of quantifying HRQOL outcomes because they can objectively document the impact of treatment. Hageman and Arrindell point out the distinction between individual and group perspectives on assessments, arguing that even groups with negligible mean changes in HRQOL scores are likely to contain individual patients whose improvement is substantial. To quantify individual change, they recommend several different statistical techniques. In the literature, several approaches have been developed to identify patients with clinically significant changes.

2.4.4.1 Effect size (ES)

Distribution-based methods to evaluate intra-individual change are evaluated in relation to sample variability. An effect size representing a standardized measure of change over time is calculated by dividing the difference between baseline and follow-up scores by the standard deviation (SD) at baseline. The effect size thus represents individual change in terms of the number of baseline SDs. Cohen has provided benchmarks that serve to guide the interpretation of effect sizes: 0.20 for "small" effects, 0.50 for "moderate" effects, and 0.80 for "large" effects. One of the limitations of this approach is that characteristics of the baseline distribution may strongly influence the effect size.
same amount of individual change produces different effect sizes depending on the sample variability at baseline: the larger the standard deviation, the smaller the effect size. In addition, the variability of change is ignored in effect size. 29

2.4.4.2 Standard response mean (SRM)

Similar to effect size is the standard response mean. As with ES, certain SRM values have been proposed to represent small, moderate, and large changes. Similar to ES, SRM is mathematically calculated based on the assumption that the change and baseline scores are normally distributed. SRM is the ratio of individual change to the SD of the sample changes. 33 A large SRM indicates that the individual change is large relative to the variability of the measurements. SRM does not ignore the variation in the change. SRM varies as a function of the effectiveness of THA. 29 The limitation of SRM is that SRM values of comparable individual changes vary with the sample: the larger the variability of the change in the sample, the smaller the SRM. Another limitation is that SRM varies as a function of the effectiveness of THA. 29

Guyatt et al. designed a responsiveness statistic associated with the standard deviation of HRQOL in a stable group. 34 The limitation of this method is that disease-specific HRQOL is not always available in a stable group.
2.4.5 Measuring individual changes

From both the patient’s and the clinician’s point of view, it would be useful to describe the changes from pre- to post operation individually, based on individual measurements instead of on sample variances. This correlates with the concerns on patient satisfaction after surgery, which is dependent on an individual’s perspective only. There are three methods that can be used. Absolute change describes the difference between follow-up (FU) and baseline (BS) HRQOL scores: FU - BS. Relative change describes the FU score in relation to BS score: FU/BS. Adjusted relative change describes the difference between FU and BS scores in relation to BS: (FU - BS)/BS. Absolute change is most commonly used in the orthopaedic literature. The significance of the change is tested by t-statistics. The limitation of this approach is that the measure has ceiling and floor effects.

2.4.5.1 Ceiling and floor effects

A ceiling effect occurs when a patient presents with a best or near best score and can improve only minimally or not at all. There is no room to detect improvement in this situation. A floor effect occurs when a patient has the worst or near worst possible score at baseline and therefore cannot experience a decrement in score, even if the patient has clinically deteriorated. In this situation, there is no room to detect deterioration. A clinically useful measure should not demonstrate ceiling or floor effects, however, every available method has a significant ceiling or floor effect, so it is necessary to find a method to measure change by adjusting the baseline measurement.
2.4.6 Satisfactory change from the patient’s perspective

The requirement for a method that distinguishes between satisfactory and unsatisfactory improvement in postoperative quality of life makes the question of the effectiveness of THA more complicated. When the aim of comparing treatments is to demonstrate a marginal, but significant difference, the research must define in advance a meaningful clinical change in patient status. A consensus on the design of clinical drug trials determined a 9.3 point change in WOMAC functional scores as a minimal clinical improvement.\(^{36}\) This change was too small to be applied to the outcomes of THA which typically show a 60-100% improvement over baseline.\(^ {11,37}\) The expected change in WOMAC functional scores after THA is four times larger than the minimal clinically important difference derived from drug trials in OA. For treatment as successful as THA, marginal differences in improvement are inadequate for documenting the positive impacts of treatment. Cohen’s criteria based on effect size is set up arbitrarily by his experience in social science and does not suit the nature of orthopaedic surgery. Because primary THA is a very effective procedure, the majority of patients in our study are considered to experience a “large effect” (>0.8) using both ES and SRM measurements. Liang et al.’s report indicates that the average responses (SRMs) to total joint replacement 12 months postoperatively are 1.05 (pain), 0.91 (mobility), and 0.55 (social function).\(^ {28}\) Therefore, instead of using a minimal clinically important difference to define a successful outcome of THA, our need is rather to differentiate between those who are able to benefit fully from the treatment and those who are not.
2.4.7 Gap in knowledge

Currently there is no common method available to differentiate patients who fully benefit from THA and those who do not, using HRQOL as the endpoint. One goal of this research is to explore a statistical technique to identify the threshold for individual change in HRQOL in order to compare inter-individual changes in relation to waiting times.
2.5 Waiting time

2.5.1 Access to the operation

The annual number of THAs increased by 25% between 1994 and 1999 in Canada\textsuperscript{38,39} It is approaching 20,000 operations nationally, with 2,500 in British Columbia (BC). Although there are geographic variations, the increase in rates does not seem to be a result of over-utilization of the procedure. On the contrary, population-based evidence gathered by Hawker et al. suggests that there is an unmet need for hip replacement.\textsuperscript{40,41} With increased average life expectancy, it is likely that the demand for hip replacement because of OA will continue to grow.

Waiting time for THA is the longest among elective surgical procedures across the country.\textsuperscript{42} Waits for hip replacement are two to three times longer than for other elective surgeries. According to figures from the BC Ministry of Health, the mean waiting time for hip replacement (primary and revisions) grew from 11 weeks in June, 1998 to 21 weeks in May, 2003 in the province. There are 2,503 patients still waiting for surgery as of May 31, 2003.\textsuperscript{43} At Vancouver Hospital, the mean wait for primary THA was 10 months in 2003 and 6 months for revision THA. The consequences of treatment delays with regard to postoperative outcomes are unclear.

Ho et al.\textsuperscript{44} surveyed 145 patients in Ontario regarding their acceptance of waiting time for joint replacement. One hundred twenty-seven responded for knee replacement. The
mean and median acceptable waiting times were 13 and 8 weeks respectively, while the mean and median unacceptable waiting times were 34 and 32 weeks (p <0.001).

2.5.2 Management of waiting lists

Waiting lists are a common tool for managing access to elective surgery. Implicit in treatment delays is the risk that health status may worsen while patients are awaiting treatment. However, little evidence is available on the health impact of delaying surgery. Queuing patients according to the urgency of treatment is generally perceived as a method to facilitate access to care within a clinically appropriate timeframe. Priority ranking algorithms have been introduced by health authorities in Scandinavian countries, New Zealand, Ontario, and most recently by the Western Canada Wait List Project (WCWL). Methods of priority ranking are based either on patient self-rated health status or on expert opinion about the benefits of an intervention for a given disease severity.

2.5.3 Deterioration while waiting

When treatment is delayed, the condition of a patient on a surgical wait list may deteriorate. Several studies investigated the effect of waiting on preoperative status. Kelly et al. used the WOMAC to measure pain levels and functional impairment at the time of consultation and prior to surgery. To see if pain and function changed while waiting, the time spent waiting was divided into 3 groups: 1-3 months, 3-6 months, and >6 months.
Compared to the 1-3 month group, the odds ratio for intensified pain was 2.0 for the 3-6 month group and 2.7 for the >6 month group. Similar results were reported with functional decline. The observed differences were not statistically significant. The authors suggest that the average wait of 4.5 months was not long enough to see a change in patient status and recommend a study on the effects of varying waiting time on the outcomes of surgery. Mahon et al. found patients who waited over 6 months for THA had a decline in function during waiting. For patients with long waits, WOMAC total scores increased by more than 10%, while walking distance decreased by more than 30 m from the time of referral to surgery.

### 2.5.4 Preoperative status and treatment outcomes

Recent studies have focused on the effect of preoperative health status on postoperative outcomes. MacWilliam et al. found that preoperative status is an important predictor of outcomes after THA. Fortin et al. examined the relationship between preoperative functional status and postoperative outcomes in a prospective cohort study using WOMAC and SF-36 measures and found that poorer preoperative function was the strongest predictor of the outcomes 6 and 24 months after THA. The authors concluded that surgery performed later in the natural history of functional decline results in worse postoperative functional status. They also noted that function and pain in patients with lower preoperative function did not improve after the operation to the level achieved by those with higher preoperative scores. Patients were pooled in this study from Boston and Montreal. Poorer baseline scores were attributed to slower access to surgery in Canada, and the question was raised as to the association between waiting time
and outcome; that is, whether waiting in the queue leads to deterioration such that final functional status is impaired. The authors were unable to establish this link since functional scores were measured only immediately preoperatively.

2.5.5 Waiting time and treatment outcomes

Several studies have examined the effects of waiting for hip and knee arthroplasty on postoperative functions. Williams et al. reported on 209 patients who underwent hip and knee replacement and were assessed before and after surgery by WOMAC and SF-36. Following surgery, all patients showed large improvements in WOMAC pain, stiffness, and physical function dimensions and marked pain relief and functional gain as measured by the SF-36. Regression analysis indicated that wait time was not a predictor of postoperative pain or functional outcome. The main shortcoming of this study is that the waiting time was recalled by patients as much as 2 years after the surgery.

Mahon et al. investigated the effects of waiting on 99 patients who underwent THA. No difference in WOMAC outcomes at 4 months was attributable to differences in waiting time. However, the authors found a larger improvement in WOMAC scores in patients who waited less than 6 months than those who waited longer. Hajat et al. conducted a large prospective cohort study looking at 7,151 patients immediately pre-surgery and at 12 months post surgery. The study showed a 4-point difference in postoperative Oxford Hip Score (range 12-60) between patients waiting more than one year and those with shorter waits. There was a wide variation in the extent of pain and disability experienced before the operation. Worse preoperative scores were associated with increased waiting time. Nilsdotter and Lohmander found that a 3-month difference in waiting time did not
result in a difference in postoperative outcomes in 124 patients who underwent THA for OA.\textsuperscript{57}

2.5.6 Gap in knowledge

Controversy exists as to the effect the queue has on function and pain while waiting, and the effect waiting has on long-term outcomes. To date, observational studies have given conflicting results. While previous studies give information only on the effects of waiting at the group level, this research focuses on individual patterns of change in HRQOL outcomes before and after THA in order to assess the effect of waiting on outcomes.
Table 1: Summary table of literature (Note: The following studies are all prospective cohort studies.)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>QOL instrument</th>
<th>Measurement</th>
<th>Method</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortin et al. (53)</td>
<td>1999</td>
<td>WOMAC SF-36</td>
<td>Preoperatively, 3, 6 months postoperatively</td>
<td>t-test, multiple regression model</td>
<td>Preoperative status is associated with treatment outcome</td>
</tr>
<tr>
<td>Fortin et al. (54)</td>
<td>2002</td>
<td>WOMAC SF-36</td>
<td>Preoperatively, 3, 6, 24 months postoperatively</td>
<td>t-test, multiple regression model</td>
<td>Preoperative status is associated with treatment outcome</td>
</tr>
<tr>
<td>MacWilliam et al. (18)</td>
<td>1996</td>
<td>HSQ(Health Status)</td>
<td>Preoperatively, 1.5, 3, 6, 12, 24 months postoperatively</td>
<td>Multiple linear regression model</td>
<td>Preoperative function is an important predictor for outcome</td>
</tr>
<tr>
<td>Kelly et al. (51)</td>
<td>2001</td>
<td>WOMAC SF-36</td>
<td>Time entering the wait list, preoperatively</td>
<td>t-test, ANOVA, logistic regression</td>
<td>Pain and function deterioration during waiting</td>
</tr>
<tr>
<td>Mahon et al. (52)</td>
<td>2002</td>
<td>WOMAC</td>
<td>Time entering the wait list, every 3-6 months until 3 months postoperatively</td>
<td>t-test</td>
<td>1. Clinically significant losses in QOL during waiting time, 2. Waiting time is not associated with outcome</td>
</tr>
<tr>
<td>Williams et al.(55)</td>
<td>1997</td>
<td>WOMAC SF-36</td>
<td>Preoperatively, 6-11 months postoperatively</td>
<td>Chi-square test, t-test</td>
<td>Waiting time is not negatively associated with outcome</td>
</tr>
<tr>
<td>Hajat et al. (56)</td>
<td>2002</td>
<td>Oxford Hip score</td>
<td>Preoperatively, 3, 12 months postoperatively</td>
<td>Multivariate linear regression model</td>
<td>Longer waiting time is associated with worse postoperative outcome</td>
</tr>
<tr>
<td>Nilsdotter et al. (57)</td>
<td>2002</td>
<td>WOMAC SF-36</td>
<td>Preoperatively, 3, 6, 12 months postoperatively</td>
<td>t-test</td>
<td>3 months difference in waiting time is not associated with worst postoperative outcome</td>
</tr>
</tbody>
</table>
3 METHODS

3.1 Overview

This study was conducted at the Vancouver Hospital & Health Sciences Centre. Ethical approval was issued by the University of British Columbia Clinical Ethics Review Board.

3.2 Study Design

This was a prospective cohort study of patients registered on the wait list for THA between March, 2001 and May, 2003 with follow-up ending in March, 2004.

3.2.1 Inclusion criteria

Patients presenting during this period at the Division of Reconstructive Orthopedics at Vancouver Hospital (VH) with a diagnosis of osteoarthritis (OA) and requiring primary total hip arthroplasty (THA) are included in the study. OA is defined by the American College of Rheumatology’s (ACR) clinical classification criteria for OA of the hip:\(^{58}\)

Pain in the hip plus two of the following:

1. ESR < 20 mm/hour
2. Radiographic femoral and/or acetabular osteophytes
3. Radiographic joint space narrowing

3.2.2 Exclusion criteria

1. Patients with previous THA to the index joint
2. Inflammatory arthritis
3. Bilateral THA performed simultaneously
4. Inability to respond to a questionnaire in English

5. Urgent surgery performed within 28 days after the decision for THA

3.2.3 Access to surgery

All patients meeting the criteria were selected from the Division of Reconstructive Orthopaedics. There are four surgeons who performed primary THA in the Division. Each surgeon has an independent waiting list and two days of scheduled operating time per week.

Fig. 1 schematizes patient flow from referral to surgery. Access is generally managed using the following steps: (1) a referring physician advises their patients to seek consultation with an orthopedic surgeon, (2) at consultation the surgeon assesses the severity of disease and the need for joint replacement, (3) if a decision is taken for surgery, the patient is registered on the surgeon’s waiting list, and (4) the patient is booked for the next available surgical appointment. There is no priority ranking system: patients are ordered based on clinical judgment with urgent cases moving ahead of the rest. There are neither systematic means, nor consistent management, of wait lists among surgeons. Each surgeon’s office operates its wait list independently. Patients are removed from the list when surgery is completed.
Figure 1: Schema of patient flow from referral to surgery

referral from community physician

orthopaedic surgery consultation

no joint replacement

watchful waiting

joint requires replacement (baseline measurement)

book surgery

surgery

one year follow-up measurement
3.2.4 Waiting time

Waiting time is defined as the interval from the mutual decision between the patient and surgeon to proceed with surgery and the time of the operation. This is the definition recommended in the ICES Report, in the WCWL, and it is widely accepted in the literature. Thus, each patient in this study had their waiting time calculated as the number of months from registration on the wait list to surgery.

3.2.5 Health-related quality of life assessment

Every patient requiring hip replacement was requested to complete the WOMAC questionnaire on the date of consultation as part of the medical offices’ administrative data collection. The questionnaire is self-administered. The medical office assistants hand each patient a WOMAC questionnaire once the decision is reached to enter the wait list for primary THA. To assess postoperative outcomes, WOMAC questionnaires were mailed at 12 months following surgery.

3.2.6 Demographic information collection

Patient’s names and provincial health numbers were used to obtain the patients’ age and gender through the medical office administrative database. Co-morbidity information was obtained through medical chart reviews measured using Charnley classification, which stratifies patients by laterality and the degree of co-morbidity, and allows meaningful comparison between groups. This is a method designed in orthopaedic research to assess co-morbidity.
Co-morbidity is classified using the Charnley classification, which includes the following categories:

- **A**: Single hip with osteoarthritis
- **B1**: Bilateral hip arthritis
- **B2**: Previous THA on contra-lateral hip
- **C**: Multiple joints arthritis or a chronic disease that affects HRQOL (specifically walking ability)

### 3.2.7 Data coding

The WOMAC questionnaire has 24 questions, with each question correlated to a Likert scale response from 0 (best health state) to 4 (worst health state). The WOMAC score for each subscale is calculated as the sum of the scores of each question included in the subscale. The range of each subscale is as follows:

- **Function**: 0-68
- **Pain**: 0-20
- **Stiffness**: 0-8

Age and gender were recorded at the time of consultation.

### 3.2.8 Data preparation

Missing data existed in response to the WOMAC questionnaire. The WOMAC function subscale (17 questions) was counted as missing when there were more than three
questions missing. Otherwise the missing question was calculated as the mean of the rest of the questions within the function subscale. The WOMAC pain subscale (5 questions) was counted as missing when there was more than one question missing. Otherwise the missing question was calculated as the mean of the rest of the questions within the pain subscale. The WOMAC stiffness subscale (2 questions) was counted as missing when both questions were missing. Otherwise, the missing question was the same as the other response.

3.3 Statistical analysis

3.3.1 Non-responder evaluation

Age in the groups of non-responders and responders was compared using the t-test. The chi-square test was used to compare non-responders and responders by gender, co-morbidity, and waiting time.

3.3.2 Summary descriptive statistics

Summary statistics were calculated using the SAS 8.1. Age, gender, recruitment time, and co-morbidity were the variables used. The frequency function and percentages were counted for those variables. Means and standard deviations of WOMAC scores were calculated, and the mean, range, and frequency of waiting times determined.
3.3.3 Responsiveness statistics

Effect size (ES) and standard response mean (SRM) for within-individual change were calculated in this study as:

\[
ES = \frac{\text{Change between follow-up and baseline WOMAC}}{(\text{standard deviation of baseline WOMAC})};
\]

\[
SRM = \frac{\text{Change between follow-up and baseline WOMAC}}{(\text{standard deviation of change})}.
\]

Means of ES and SRM were also described in the study.

3.3.4 Log-linear regression (Model 1)

Since the distribution of follow-up scores is skewed, one cannot use it in the linear regression analysis as an outcome variable. Therefore, the following association was studied:

\[
Model 1. \log(FU) = \alpha + \theta \cdot BS + \sigma \cdot \epsilon,
\]

where \(FU\) is follow-up WOMAC score, \(BS\) is baseline WOMAC score, the error term \(\epsilon\) follows a normal distribution with a mean of 0 and a standard deviation of 1, and \(\sigma\) is a fixed constant that changes the location of the expected value. (See Section 4.5.5)

3.3.4.1 Parameter estimation

Parameters in this model can be estimated by the ordinary least squares (OLS) method. The method produces the best linear unbiased estimates of the regression coefficients, regardless of the shape of the distribution of \(\epsilon\). When \(\epsilon\) is assumed to be normally
distributed, the OLS estimates will also be maximum likelihood estimates and will have minimum variance among all possible estimators.

Age, gender, co-morbidity, and waiting time are considered possible confounders that need to be controlled. Each variable was run separately with the baseline WOMAC score in the log-linear regression model to test its possible role as a confounder.

3.3.4.2 Data censoring

The lowest score for WOMAC function, pain and stiffness is 0, corresponding to a perfect health state. It is unlikely that patients before and after such a major operation as THA would be perfectly healthy. Therefore, as a measurement tool, the WOMAC questionnaire has a limited ability to provide information on HRQOL for patients at the lowest extreme of the scale, that is, below a score of 1.

Since a true score is unknown when the score reading is between 0 and 1, we regard measurements below 1 as left-censored observations. The Tobit model was used to incorporate the left-censored observations in the regression analysis, and the maximum likelihood method to estimate the probabilities of log(FU) given the baseline WOMAC score. The regression analysis was conducted by running the SAS 8.1 PROC LIFEREG procedure. We transformed the observed follow-up score as follows:

\[
FU = 0.9, \text{ if } FU \leq 0.9 \\
FU = FU, \text{ if } FU > 0.9
\]
SAS codes used in the analysis were as follows:

```sas
data w;
set wosfox;
if wf=0 then wf=0.9;
else wf=wf;
run;

data w1;
set w;
if wf=0.9 then lower=.;
else lower=wf;
label bwf = 'Baseline Score';
label wf = 'Follow-up Score';
run;

title 'FU Vs. BS WOMAC function score on logarithm scale';
proc lifereg data=w1 outest=OUTEST(keep=_scale_);
model (lower,wf)=bwf/d-lnormal;
output out=OUT xbeta=Xbeta cdf=prob p=predwf std=std;
run;
quit;
```

### 3.3.4.3 Goodness of fit statistics for the regression model

A linear regression model was used with the dependent variable log(FU) and we obtained the log likelihood for this model. In the final analysis of the model with selected predictors, we used

\[-2 \times [\text{log likelihood in final model} - \text{log likelihood in reference model}]\]

as the test statistics value. Because this value follows a chi-square distribution, the p-values are found in a chi-square table.

### 3.3.4.4 Criteria for individual change

Using the regression estimates derived from Model 1, the median and confidence intervals (CI) of the follow-up WOMAC scores for a certain baseline WOMAC score can be estimated.
Using the lower 90% CI as the criteria, study patients were divided into two outcome groups (See Fig. 8):

1. "Better than expected" outcome: Those who achieved a follow-up score lower than the lower 90% CI.

2. "Not better than expected" outcome: Those who achieved a follow-up score higher than the upper 90% CI.

3.3.5 Chi-square test

Chi-square tests were performed to evaluate the association between waiting time and outcomes. Both waiting time and outcomes were treated as binary variables, as described above. The percentage of those achieving “better than expected” outcomes in the short waiting group were compared with those achieving “better than expected” outcomes in the long waiting group. P-values were reported in a 2-tailed manner.

3.3.6 Multivariate logistic regression (Models 2 and 3)

Multivariate logistic regression models were used to examine the association between waiting time and the probability of “better than expected” outcomes and to measure the size of the effect. Waiting time was studied both as a continuous and binary variable.

To create a binary independent variable, waiting time was dichotomized as follows:

- Short waiting time: waited less than or equal to 6 months
- Long waiting time: waited over 6 months.
For binary waiting time we use the following model

\[ \text{Model 2. } \logit(\text{outcome}) = a + \beta LW + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n, \]

where \( LW \) is 1 when the wait time is over 6 months, 0 otherwise. In this case, the estimate of \( \beta \) gives the log-odds ratio associated with longer times.

For continuous waiting time we used the following model

\[ \text{Model 3. } \logit(\text{outcome}) = a + \beta (\text{waiting time}) + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n, \]

where the outcome is 1 when the outcome is “better than expected” and where 0 is “not better than expected”; \( \logit = \log[p/(1- p)] \), \( p = \Pr(\text{outcome} = 1) \). The effect size is measured by the estimate of \( \beta \), which shows the change in the log-odds ratio for a 1 month increase in wait time. Waiting time is the variable of interest; the other variables \( X_1, X_2, \ldots, X_n \) were entered into the model to obtain adjusted estimates.

There were four variables in the final models (Model 2 and 3). The p-value (2 sided) in the regression provided the statistical significance for the association between outcomes and the variable of interest.

### 3.3.6.1 Variable selection

Variables were selected in this regression model using the SAS 8.0 PROC LOGISTIC procedure’s stepwise selection method. Those variables having a p value < 0.1 were retained. Waiting time, age, gender, and co-morbidity were used in the selection procedure.
3.3.6.2 Confounders

Although there are variables not selected in the model, demographic factors and co-morbidity could be potential confounders. Each variable was separately tested with the variable of interest (waiting time) by logistic regression. If the regression coefficient for waiting time changed in the model by adding or removing the variable, then the variable was put in the regression model to adjust for its confounding effect.

3.3.6.3 Interaction

Interactions between waiting time, demographic factors, and co-morbidity were tested by putting the interaction term in the regression model. The interaction term is not significant when the p-value is bigger than 0.05.

3.3.6.4 Goodness of fit test

The LACKFIT function in the PROC LOGISTIC procedure was used to perform the Hosmer and Lemeshow goodness-of-fit test. The discrepancies between the observed and expected number of observations in these groups are summarized by the Pearson chi-square statistic, which is then compared to a chi-square distribution with \( t \) degrees of freedom, where \( t \) is the number of groups minus 2. A small p-value suggests that the fitted model is not an adequate model.
3.3.6.5 Model checking

Pearson's residuals were used to identify observations that are not explained well by the model. The PROC LOGISTIC procedure in SAS provides the Pearson's residual plot for the regression model. This plot should show no particular pattern.
4 RESULTS

4.1 Study population

Between March, 2001 and May, 2003, 668 patients with a diagnosis of OA were entered in the waiting list for total hip replacement at Vancouver Hospital. Two hundred thirteen patients completed a baseline self-assessment in the medical office at the time of their consultation. Twelve patients were excluded because they had a previous hip arthroplasty of the index joint or were having both hips replaced at the same surgery. That left 201 baseline assessments in the study. One year after their surgery, 147 patients returned a follow-up WOMAC questionnaire by mail.
4.2 Waiting time

Fig. 2 shows the distribution of waiting times in this study. The range of waiting times in this sample population is 1 to 18 months. The mean waiting time is 6.3 months; the standard deviation is 4.4 months. There are 77 patients who waited less than or equal to 6 months, and 70 patients who waited for more than 6 months.

Figure 2: The distribution of waiting time (in months)
4.3 Patient characteristics

Table 2 shows the distribution of patients by demographic characteristics. Of 147 patients in the study population, the average age is 64.8 years. There were 18 patients (12%) younger than 50; 33 patients (22%) between 50 and 59; 42 patients (29%) between 60 and 69; 38 patients (26%) between 70 and 79; and 16 patients (11%) over 80 years old. There were 83 females (56%) and 66 males (44%) in the study. The study period lasted from 2001 until 2003. Eighty-five patients (55%) were enrolled in 2001, 52 patients (35%) in 2002, and 15 patients (10%) in 2003. Seventy-two patients (50%) had only one joint involved with OA; 34 patients had bilateral disease. Of these 34 patients, there were 18 with contra-lateral hip replacement prior to the index surgery and 16 patients with moderate to severe OA in the contra-lateral hip. Thirty-nine patients (27%) had multiple joints involved with OA or had a chronic systematic disease.
Table 2: Patient demographic statistics

<table>
<thead>
<tr>
<th>Age</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-49</td>
<td>18(12)</td>
</tr>
<tr>
<td>50-59</td>
<td>33(22)</td>
</tr>
<tr>
<td>60-69</td>
<td>42(29)</td>
</tr>
<tr>
<td>70-79</td>
<td>38(26)</td>
</tr>
<tr>
<td>80+</td>
<td>16(11)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>64(44)</td>
</tr>
<tr>
<td>Female</td>
<td>83(56)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enrollment Period</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>80(55)</td>
</tr>
<tr>
<td>2002</td>
<td>52(35)</td>
</tr>
<tr>
<td>2003</td>
<td>15(10)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co-morbidity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>72(50)</td>
</tr>
<tr>
<td>B1</td>
<td>16(11)</td>
</tr>
<tr>
<td>B2</td>
<td>18(12)</td>
</tr>
<tr>
<td>C</td>
<td>39(27)</td>
</tr>
</tbody>
</table>
4.4 Non-responder and responder comparison

Two hundred and one patients completed a baseline assessment; 147 patients (76%) responded to our one-year follow-up assessment. The non-responders were compared to the responders with respect to age, gender, and co-morbidity.

Table 3 shows the difference between these two groups. The mean age of responders was 65 years. This is significantly higher than the mean age (62 years) of the non-responders (p=0.001). There was a greater proportion of non-responders in the short waiting group, but the difference is not statistically significant (p=0.444). There were 44% male patients among the responders and only 32% among the non-responders; the difference, however, is not statistically significant (p=0.097). Forty-nine percent of patients in the Charnley A class responded to the follow-up assessment compared to 43% among the non-responders; the difference was not statistically significant (p=0.793).
Table 3: Comparison between responders and non-responders

<table>
<thead>
<tr>
<th></th>
<th>Non-responders</th>
<th>Responders</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years</td>
<td>62</td>
<td>65</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Waiting time</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short waits</td>
<td>25(46%)</td>
<td>77(51%)</td>
<td></td>
</tr>
<tr>
<td>Long Waits</td>
<td>29(54%)</td>
<td>70(49%)</td>
<td>0.444</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>36(68%)</td>
<td>83(56%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17(32%)</td>
<td>64(44%)</td>
<td>0.097</td>
</tr>
<tr>
<td><strong>Co-morbidity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charnley A</td>
<td>23(43%)</td>
<td>72(49%)</td>
<td></td>
</tr>
<tr>
<td>Charnley B1</td>
<td>8(15%)</td>
<td>16(11%)</td>
<td></td>
</tr>
<tr>
<td>Charnley B2</td>
<td>6(11%)</td>
<td>18(12%)</td>
<td></td>
</tr>
<tr>
<td>Charnley C</td>
<td>16(30%)</td>
<td>39(27%)</td>
<td>0.793</td>
</tr>
</tbody>
</table>
4.5 WOMAC function

4.5.1 Distribution of baseline and follow-up WOMAC function scores

Fig. 3 displays the distribution of baseline WOMAC functional scores (scale 0-68). It follows a symmetrical distribution. The mean (SD) baseline WOMAC function score is 39 (13).

Figure 3: The distribution of baseline WOMAC function scores
Fig. 4 displays the distribution of follow-up WOMAC function scores (scale 0-68). It follows a skewed distribution because the follow-up outcome is nearly as good as a full recovery in a normal person; that is, the follow-up outcomes have a floor limit as a score of 0 (best function). The mean follow-up WOMAC function score is 14 and the standard deviation is 14. As the distribution is skewed, one cannot use the follow-up score in the linear regression analysis as an outcome variable.

Figure 4: The distribution of follow-up WOMAC function scores
4.5.2 Responsiveness statistics

Individual changes in WOMAC function scores between baseline and follow-up measurements were evaluated by two statistics: the effect size (ES) and the standard response mean (SRM). Since the responsiveness to THA was so high, the majority of patients experienced a large effect by Cohen’s criteria, i.e. >0.8. Seventy-eight percent of patients had a large change as measured by ES, with a mean ES of 1.93. Seventy-five percent of patients had a large SRM (>0.8), with a mean SRM of 1.39. The variation of individual change was also substantial as measured by the coefficient of variation (65%). In studies on outcomes of treatment as successful as THA, even groups with a large mean change in HRQOL scores are likely to contain individual patients whose improvement is less than expected.

The goal of this study is to find out why some patients benefit fully from treatment while others do not. In particular, the effects of waiting time on the probability of better than expected outcomes were investigated. Therefore, instead of using an arbitrary criterion for change to define a successful outcome of THA, one needs to determine the expected follow-up scores based on baseline information.
4.5.3 Prediction of outcomes based on individual baseline QOL

Many studies have shown that pre-surgery HRQOL is a strong predictor of post-surgery HRQOL outcomes. Fig. 5 displays the plot between baseline WOMAC function scores and postoperative WOMAC function scores. It shows that baseline WOMAC function is a strong positive predictor only when the patient has a relatively better baseline score. (Note: the lower the score, the better the HRQOL). The effect is restricted by the floor effect.

Figure 5: Baseline WOMAC function scores and follow-up WOMAC function scores
4.5.4 Logarithm of follow-up WOMAC function scores

Follow-up WOMAC function scores follow a skewed distribution in this study. However, the logarithms of the follow-up WOMAC function scores follow a symmetrical distribution as shown in Fig. 6.

Figure 6: The distribution of the logarithms of follow-up WOMAC function scores
4.5.5 Parameter estimation

As a result of the skewed distribution, the following association was studied:

\[
\log(FU) = \alpha + \beta BS + \sigma e,
\]

where \(FU\) is the follow-up WOMAC score and \(BS\) is the baseline WOMAC score. The error term \(e\) follows a normal distribution with a mean of 0 and a standard deviation of 1, and \(\sigma\) is a fixed constant that changes the location of the expected value.

Table 4 shows the parameter estimates obtained through this regression analysis.

**Table 4: Parameter estimation for the log-linear regression**

<table>
<thead>
<tr>
<th></th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.34</td>
</tr>
<tr>
<td>Beta</td>
<td>0.03</td>
</tr>
<tr>
<td>Scale</td>
<td>1.23</td>
</tr>
</tbody>
</table>
For the sample population, the estimate of the expected value of the lognormal distribution is given by:

\[ \log(FU) = 1.34 + 0.03 \times BS. \]

Based on this regression, baseline WOMAC function score is a significant predictor for the follow-up WOMAC function score (p=0.0005).

Since age is an important demographic risk factor for HRQOL, the regression estimates were adjusted for age, and we found that the regression coefficient for baseline function score is only slightly changed after the adjustment. Age did not have a significant effect on follow-up WOMAC function score (p=0.1693). Thus, the effect of baseline function on postoperative function is independent of age.

Using the log-linear regression model, gender, co-morbidity, and waiting time were investigated as covariates. None of the variables listed above were significant predictors for follow-up WOMAC function scores. There was no difference in the regression coefficient for baseline WOMAC function scores with or without these covariates.

### 4.5.6 Goodness of fit statistics for the Tobit model

The log likelihood in this model is -219, the reference model has a log likelihood of -231, and the test statistic for the goodness of fit is \(-2 \times [-219 - (-231)] = 24\). The p-value for the chi-square distribution is less than 0.001; therefore the null hypothesis of no association between BS and Log (FU) WOMAC function score is rejected.
4.5.7 Criteria for function outcomes

Through this regression analysis an expected postoperative outcome for each baseline WOMAC function score was obtained. The regression line shown in Fig. 7 is the estimation of follow-up WOMAC function scores based on specific baseline scores. Fig. 8 shows the median score and the 90% confidence interval for the regression line in relation to baseline scores.

Figure 7: The regression line for the logarithm of follow-up score versus baseline WOMAC function

**FU function score on logarithm scale Vs. BS WOMAC function score**

![Graph showing regression line and quantile values](image)

*Note: Red dots are the average follow-up WOMAC function scores when categorized by baseline WOMAC scores into categories by 5.*
* The red line is the lower 90% confidence interval for the median of expected outcomes. The lower the score, the better the QOL.

Using the lower 90% confidence interval as a cut off point associated with the baseline score, the study patients were divided into two groups:

- Patients below the line were considered to have achieved a "better than expected" outcome
- Patients above the line were considered to have achieved a "not better than expected" outcome
4.5.8 Waiting time effect

4.5.8.1 Percent achieving a “better than expected” functional outcome

Overall, 55 patients (37%) achieved a “better than expected” outcome.

Fig. 9 shows that there are 12% fewer patients with a “better than expected” functional outcome in the long waiting group than in the short waiting group (p=0.15, chi-square test).

Figure 9: The percent achieving a “better than expected” functional outcome in short and long waiting groups

| Percentage “better than expected” function outcome |
|---------------------------------|-----------|
| 50%                             | 43%       |
| 45%                             | 31%       |
| 40%                             |           |
| 35%                             |           |
| 30%                             |           |
| 25%                             |           |
| 20%                             |           |
| 15%                             |           |
| 10%                             |           |
| 5%                              |           |
| 0%                              |           |

Legend: 
- **Short waiting**
- **Long waiting**
4.5.8.2 Logistic regression results

4.5.8.2.1 Odds ratio (long wait group vs. short wait group)

Multivariate logistic regression models were used to evaluate the effect of waiting time. Waiting time was found to be a significant risk factor at the significance level of 0.1 (p=0.057). Gender was also a significant risk factor at this level (p=0.09). Age was not a significant factor (p=0.22), and neither was co-morbidity (p=0.63). Because the OR of waiting time changes in the models with age, gender and co-morbidity, possible confounders were adjusted in the logistic regression model. No significant interaction term between waiting time and other covariates was found in this model.

As shown in Table 5, the odds ratio of achieving “better than expected” outcome is 0.6 (CI 0.3-1.1) in the long wait group compared to the short wait group. After adjustment by age, gender, and co-morbidity, the OR of achieving a “better than expected” outcome is 0.5 (CI 0.2 -1.1) in the long wait group compared to the short wait group.

Table 5: Relationship between waiting time and one year outcomes measured by WOMAC function scores
*Adjusted by age, gender, and co-morbidity (p=0.057)

<table>
<thead>
<tr>
<th>Waiting time</th>
<th>OR</th>
<th>Adjusted OR *</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=6 months</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>&gt;6 months</td>
<td>0.6 (0.3-1.1)</td>
<td>0.5 (0.2 -1.1)</td>
</tr>
</tbody>
</table>
4.5.8.2.2 Model checking

The Pearson's residual plot did not show any specific patterns. The Hosmer and Lemeshow goodness-of-fit test of this logistic regression model is 0.58, which indicates that the model is adequately fitted.

Figure 10: Residuals for the logistic regression model
4.5.8.2.3 Odds ratio (waiting time as a continuous variable in month units)

Waiting time was also considered as a continuous variable in the logistic regression. Each additional month in waiting time is associated with a 7% decrease in the odds of a “better than expected” outcome, OR = 0.93 (CI 0.86-1.01). After adjustment for age, gender, and co-morbidity, the OR estimates did not change, 0.92 (CI 0.85-1.00). Waiting time is a significant risk factor for “better than expected” outcomes measured by the WOMAC function subscale (p=0.05).

4.5.8.2.4 Model checking

The Pearson’s residual plot did not show any specific patterns or outliers. The Hosmer and Lemeshow goodness-of-fit test of this logistic regression model is 0.31, which indicates that the model is adequately fitted.
4.6 WOMAC pain

4.6.1 Distribution of baseline and follow-up WOMAC pain scores

Fig. 11 displays the distribution of baseline WOMAC pain scores (scale 0-20). It follows a symmetrical distribution. The mean (SD) baseline WOMAC pain score is 11.6 (3.6).

Figure 11: The distribution of baseline WOMAC pain scores
Fig. 12 displays the distribution of follow-up WOMAC pain scores (scale 0-20). It follows a skewed distribution because the follow-up outcome is nearly as good as a full recovery in a normal person; that is, the follow-up outcome has a floor limit score of 0 (least pain). The mean follow-up WOMAC pain score is 2.6 and the standard deviation is 3.0. As the distribution is skewed, one cannot use it in the linear regression analysis as an outcome variable.

Figure 12: The distribution of follow-up WOMAC pain scores
4.6.2 Responsiveness statistics

Individual changes between baseline and follow-up pain scores were measured by two statistics: the effect size (ES) and the standard response mean (SRM). Eighty-four percent of patients have achieved a large effect (ES > 0.8). The mean ES was 0.83. Eighty percent of patients achieved a large SRM (SRM > 0.8) with a mean of 0.8. As discussed before, because the responsiveness to THA was so great, the majority of patients experienced a large effect. But the variation in outcomes is still large as measured by the coefficient of variation (46%). Therefore, the factors that would explain the difference between a “very good outcome” and a “not so good outcome” were investigated. (See section on function)
4.6.3 Prediction of outcome based on individual baseline QOL

Many studies have shown that pre-surgery HRQOL is a strong predictor of post-surgery HRQOL outcomes. Fig. 13 displays the plot of the baseline WOMAC pain scores against postoperative WOMAC pain scores. It shows that baseline WOMAC pain is a strong positive predictor only when the patient has a relatively better baseline score. (Note: the lower the score, the better the HRQOL). The effect is restricted by the floor effect.

Figure 13: Baseline WOMAC pain scores and follow-up WOMAC pain scores
4.6.4 Logarithm of follow-up WOMAC pain scores

The follow-up WOMAC pain scores follow a skewed distribution. However, the logarithms of the follow-up WOMAC pain scores follow a rather symmetrical distribution as shown in Fig. 14.

Figure 14: The distribution of the logarithms of follow-up WOMAC pain scores
4.6.5 Parameter estimation

Therefore the following association was studied:

\[ \log(FU) = \alpha + \beta BS + \sigma e \]

Where \( FU \) is the follow-up WOMAC pain score and \( BS \) is the baseline WOMAC pain score. The error term \( e \) follows a normal distribution with a mean of 0 and standard deviation of 1, and \( \sigma \) is a fixed value. The estimates for the parameters achieved through this regression model are shown in Table 6.

Table 6: Parameter estimation for the log-linear regression

<table>
<thead>
<tr>
<th></th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.66 (-1.43, 0.12)</td>
</tr>
<tr>
<td>Beta</td>
<td>0.09 (0.03, 0.15)</td>
</tr>
<tr>
<td>Scale</td>
<td>1.21 (1.03, 1.43)</td>
</tr>
</tbody>
</table>

For our sample population, the estimate of the expected value of the lognormal distribution is given by:

\[ \log (FU) = -0.06 + 0.09 BS + 1.21 \]

Baseline WOMAC pain score is a very significant predictor for the follow-up WOMAC pain score (\( p=0.0036 \)).
Because age is an important demographic risk factor for HRQOL, the regression model was adjusted for age. However, we found that age has an insignificant effect (p=0.2135) on follow-up WOMAC pain. The coefficients of baseline WOMAC pain levels were only slightly changed after adjusting for age. Thus, the effect of baseline pain on postoperative pain is independent of age in this study.

The effect of gender, co-morbidity, and waiting time were also investigated as covariates. None of the variables listed above were significant predictors for follow-up WOMAC pain score. There was no difference in the regression coefficient for baseline WOMAC pain score with or without these covariates.

4.6.6 Goodness of fit statistics for the Tobit model

The log likelihood in this model is -174, the reference model has a log likelihood of -192, and the test statistic for the goodness of fit is \(-2\times[-192-(-174)]\) = 36. The p-value for the chi-square distribution is less than 0.001; therefore the log-linear regression is well fitted.

4.6.7 Criteria for pain outcome

Through this regression, an expected postoperative outcome for each baseline WOMAC pain score was obtained. Thus, the median and 90% confidence interval of expected pain outcomes as they relates to different baseline pain scores were obtained.
Using the lower 90% confidence interval for the regression line as a cut off point, the sample population was divided into two groups:

- Patients below the line were considered to have achieved a "better than expected" outcome
- Patients above the line were considered to have achieved a "not better than expected" outcome

### 4.6.8 Waiting time effect

Using these individualized cut off points, patients were differentiated by whether they exceeded or failed to achieve the expected outcome through surgery. Fifty-five patients (37%) achieved a “better than expected” outcome and 92 patients (63%) achieved a “not better than expected” outcome.
4.6.8.1 Percent achieving a “better than expected” pain outcomes

Three percent fewer patients in the long waiting group achieved a “better than expected” pain outcome, compared to short waits, in Fig. 15. As measured by the chi-square test, the difference was not statistically significant (p=0.68, chi-square test).

Figure 15: The percent achieving a “better than expected” pain outcome in short and long wait groups
4.6.8.2 Logistic regression results

4.6.8.2.1 Odds ratio (long wait group vs. short wait group)

Because the OR of waiting time changes in the models with or without age, gender, and co-morbidity, the possible confounders were adjusted in the logistic regression model. No significant interaction term between waiting time and other covariates was found in this model. As shown in Table 7, the odds ratio of achieving “better than expected” outcomes is 0.9 (CI 0.5-1.8) in the long wait group compared to the short wait group. After adjusting for age, gender, and co-morbidity, the OR of achieving a “better than expected” outcome is 0.8 (CI 0.4-1.5) in the long wait group compared to the short wait group.

Table 7: Relationship between waiting time and one year outcomes of treatment measured by WOMAC pain scores

<table>
<thead>
<tr>
<th>Waiting time</th>
<th>OR</th>
<th>Adjusted OR*</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=6 months</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>&gt;6 months</td>
<td>0.9(0.5-1.8)</td>
<td>0.8(0.4-1.5)</td>
</tr>
</tbody>
</table>

*Adjusted by age, gender, and co-morbidity (p=0.623)
4.6.8.2.2 Model checking

The Pearson’s residual plot did not show any specific patterns or any outliers. The Hosmer and Lemeshow goodness-of-fit test of this logistic regression model is 0.49, which indicates the model is adequately fitted.

4.6.8.2.3 Odds ratio (waiting time as a continuous variable in month units)

The waiting was also entered in the logistic regression as a continuous variable. The effect of additional month of waiting time on the probability of “better than expected” outcomes measured by WOMAC pain scores was not statistically significant (p=0.34) with OR = 0.97 (CI 0.90-1.04). After adjusting for age, gender, and co-morbidity, the OR estimate did not change, 0.96 (CI 0.89-1.04).

4.6.8.2.4 Model checking

The Pearson’s residual plot did not show any specific patterns or outliers. The Hosmer and Lemeshow goodness-of-fit test of this logistic regression model is 0.13, which indicates that the model is not adequately fitted.
4.7 WOMAC stiffness

4.7.1 Distribution of baseline and follow-up WOMAC stiffness scores

Fig. 16 displays the distribution of baseline WOMAC stiffness scores (scale 0-8). It follows a rather normal distribution. The mean (SD) baseline WOMAC stiffness scores is 4.8 (1.7).

Figure 16: The distribution of baseline WOMAC stiffness scores
Fig. 17 displays the distribution of follow-up WOMAC stiffness scores (scale 0-8). It follows a skewed distribution because the follow-up outcome is nearly as good as full recovery in a normal person; that is, the follow-up outcome has a floor limit score of 0 (least stiffness). The mean follow-up WOMAC stiffness score is 2.2 and the standard deviation is 1.6. As the distribution is skewed, one cannot use it in the linear regression analysis as an outcome variable.

Figure 17: The distribution of follow-up WOMAC stiffness scores
4.7.2 Responsiveness statistics

Individual changes between baseline and follow-up function scores were measured by two statistics: the effect size (ES) and the standard response mean (SRM). Sixty-five percent of patients achieved a large effect (ES>0.8). The mean ES was 0.65. The standard response means (SRM) was calculated individually as well. Sixty-five percent of patients had a large SRM (SRM>0.8) with a mean of 0.65. The coefficient of variation in change between preoperative and postoperative WOMAC stiffness outcomes was 83%. As discussed before, the factors that would result in a difference between “very good outcome” and “not so good outcome” were investigated.
4.7.3 Logarithms of follow-up WOMAC stiffness scores

The follow-up WOMAC stiffness scores follow a skewed distribution. However, the logarithms of the follow-up WOMAC stiffness scores follow a rather symmetrical distribution as shown in Fig. 18.

Figure 18: The distribution of the logarithm of follow-up WOMAC stiffness scores
4.7.4 Parameter estimation

As a result of the skewed distribution, the following association was studied:

\[ \log(FU) = \alpha + \beta \cdot BS + \sigma \varepsilon \]

Where \( FU \) is the follow-up WOMAC stiffness score and \( BS \) is the baseline WOMAC stiffness score. The error term \( \varepsilon \) follows a normal distribution with a mean of 0 and a standard deviation of 1, and \( \sigma \) is a fixed value.

The estimates for the parameters achieved through this regression model are shown in Table 8.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.02</td>
<td>(-0.34, 0.38)</td>
</tr>
<tr>
<td>Beta</td>
<td>0.12</td>
<td>(0.06, 0.19)</td>
</tr>
<tr>
<td>Scale</td>
<td>0.69</td>
<td>(0.60, 0.79)</td>
</tr>
</tbody>
</table>

For our sample population, the estimate of the expected value of the lognormal distribution is given by:

\[ \log(FU) = 0.02 + 0.12 \cdot BS + 0.69 \]

The baseline WOMAC stiffness score is a significant predictor for the follow-up WOMAC stiffness score (\( p=0.0004 \)).
Age is an important demographic risk factor for HRQOL. Therefore the regression model was adjusted for age. We found that age has an insignificant effect \( p=0.2762 \) on the follow-up WOMAC stiffness score. The coefficients of the baseline WOMAC stiffness levels were only slightly changed after adjusting for age. Thus, the effect of baseline stiffness on postoperative stiffness is independent of age. Gender, co-morbidity, and waiting time were also investigated as covariates. None of these variables were significant predictors for follow-up WOMAC stiffness. There was no difference in the regression coefficient for baseline WOMAC stiffness with or without these covariates.

4.7.5 Goodness of fit statistics for the Tobit model

The log likelihood in this model is -146, the reference model has a log likelihood of -159, and the test statistic for the goodness of fit is \(-2\times[-159-(-146)]=26\). The p-value for the chi-square distribution is less than 0.001. Therefore the log-linear regression is well fitted.

4.7.6 Criteria for stiffness outcome

Through this regression, an expected postoperative outcome for each baseline WOMAC stiffness score was obtained. The median and 90% confidence interval of expected stiffness outcomes as they relate to different baseline stiffness scores were calculated.
Using the lower 90% confidence interval for the regression line as a cut off point, study patients were dichotomized into two groups:

- **Patients below the line were considered to have achieved a “better than expected” outcome**
- **Patients above the line were considered to have achieved a “not better than expected” outcome.**

Forty-eight patients (33%) achieved a “better than expected” outcome and 99 patients (67%) achieved a “not better than expected” outcome.
4.7.7 Waiting time effect

4.7.7.1 Percent achieving a “better than expected” stiffness outcome

Using these individualized cut off points, patients were differentiated by whether they exceeded or failed to achieve the expected outcome through surgery. Three percent fewer patients in the long waiting group achieved a “better than expected” stiffness outcome compared to short waits, in Fig. 15. As measured by the chi-square test, the difference was not statistically significant (p=0.76, chi-square test).

Figure 19: The percent achieving a “better than expected” stiffness outcome in short and long waiting groups
4.7.7.2 Logistic regression results

4.7.7.2.1 Odds ratio (long wait group vs. short wait group)

Because the OR of waiting time changes in models with or without age, gender, and co-morbidity, these possible confounders were adjusted in the logistic regression model. There was no significant interaction term found in this model. As shown in Table 9, the odds ratio of achieving a “better than expected” outcome is 0.9 (CI 0.4-1.8) in the long wait group compared to the short wait group. After adjustment by age, gender, and co-morbidity, the OR of achieving a “better than expected” outcome is 0.8 (CI 0.4-1.6) in the long wait group compared to the short wait group.

Table 9: Relationship between waiting time and one year outcome of treatment measured by WOMAC stiffness scores

<table>
<thead>
<tr>
<th>Waiting time</th>
<th>OR</th>
<th>Adjusted OR*</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=6 months</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>&gt; 6 months</td>
<td>0.9(0.4-1.8)</td>
<td>0.8(0.4-1.6)</td>
</tr>
</tbody>
</table>

*Adjusted by Age, Gender, and Co-morbidity (p=0.51)
4.7.7.2.2 Model checking

The Pearson’s residual plot did not show any specific patterns or outliers. The Hosmer and Lemeshow goodness-of-fit test of this logistic regression model is 0.46, which indicates that the model is adequately fitted.

4.7.7.2.3 Odds ratio (Waiting time as a continuous variable in month unit)

The waiting was also entered in the logistic regression as a continuous variable. The effect of an additional month of waiting time on the probability of “better than expected” outcomes measured by WOMAC pain scores was not statistically significant (p=0.66) with OR = 0.99 (CI 0.92-1.07). After adjusting for age, gender, and co-morbidity, the OR estimates did not change, 0.98 (CI 0.91-1.07).

4.7.7.2.4 Model checking

The Pearson’s residual plot did not show any specific patterns or outliers. The Hosmer and Lemeshow goodness-of-fit test of this logistic regression model is 0.09, which indicates that the model is not adequately fitted.
5 DISCUSSION AND CONCLUSION

5.1 Overview

In this chapter, I discuss the results of this study. First, I discuss the findings on the association between baseline and follow-up HRQOL scores. Next I discuss the findings on the effects of delay in access to primary total hip arthroplasty and its relationship to postoperative outcomes. Limitations of the study are also discussed. Finally, I present my conclusions and suggest directions for future studies.

5.2 Discussion

In this observational study, OA patients awaiting THA from the time of surgical consultation were prospectively followed. One year after THA, the majority of patients experienced substantial improvements in their physical function as measured by the WOMAC self-assessment instrument. We observed, as have other investigators, that there is a large variation, measured by the coefficient of variance in change, between preoperative score and postoperative outcomes.

5.2.1 Association between baseline and follow-up WOMAC scores

In studies on outcomes of treatment as successful as THA, even groups with a large mean change in HRQOL scores will contain individual patients whose improvement is less than expected. Our goal was to find out why some patients benefit fully from treatment while
others do not. As a first step, a linear regression model for the average logarithm of follow-up scores was developed to examine the association between baseline and follow-up WOMAC scores.

WOMAC function, pain, and stiffness follow-up scores were found to be strongly associated with their baseline scores. The effects of age, gender, and co-morbidity on follow-up WOMAC scores were not statistically significant.

5.2.1.1 Utilization of the method

Total hip arthroplasty (THA) is a highly effective treatment. However, patients do not benefit equally. It is important to consider the potential advantage attained by those patients who benefit most from the treatment as a goal to be aimed at for all patients. The linear regression model provides a means of understanding and communicating the relationship between baseline and expected outcomes. It provides an intuitively satisfying measure of HRQOL based on individual scores before surgery. It is also a means to investigate the effect of other covariates on expected outcomes. Surgeons in the clinical setting will find it useful for counseling patients about what they can expect to achieve with the surgery.
5.2.2 Effect of waiting on WOMAC outcomes of THA

5.2.2.1 WOMAC function

We found that there was a strong association between waiting time and the odds of a “better than expected” functional outcome. The significant decrease (12%) in the proportion achieving a “better than expected” functional outcome among the long waiters compared to the short waiters indicates that longer waiting may reduce the chance of maximizing surgical benefits. The odds of a “better than expected” functional outcome decrease by 8% for each additional month on the wait list. Compared to the short wait group (wait <= 6 months), the odds of a “better than expected” functional outcome were 50% lower for the long wait group (wait > 6 months).

5.2.2.2 WOMAC pain

This study found that there is no statistically significant association between waiting time and the odds of achieving a “better than expected” pain outcome. The decrease (3%) in the proportion of those achieving a “better than expected” pain outcome among the long-waiters compared to the short waiters indicates that longer waiting may reduce the chance of a “better than expected” pain reduction after THA. The odds of achieving a “better than expected” pain outcome are 80% lower for a patient in the long wait group (wait > 6 months) compared to a patient in the short wait group (wait <= 6 months). However, the effect is not statistically significant.
5.2.2.3 WOMAC stiffness

This study showed that there is no association between waiting time and the odds of achieving a “better than expected” stiffness outcome. The decrease (3%) in the proportion of those achieving a “better than expected” stiffness outcome among the long-waiters compared to the short waiters indicates that longer waiting may reduce the chance of “better than expected” stiffness reduction after surgery. The odds of achieving a “better than expected” stiffness outcome are 80% lower for a patient who waited in the long wait group (wait > 6 months) compared to a patient in the short wait group (wait <= 6 months). However, the effect is neither statistically, nor clinically, significant.

5.2.2.4 Cofounders and interactions

This study found that age, gender, and co-morbidity were confounders for the effect of waiting time on the probability of “better than expected” outcomes. The finding has been taken into consideration and adjustments made for the confounders in the final model. No interaction terms of waiting time and the variables listed above are significant variables in the model.

5.2.3 External validity

Vancouver Hospital (VH) is a tertiary referral center and teaching hospital for the University of British Columbia (UBC). The demographics of arthroplasty patients, however, are not different from elsewhere. In their population-based study, Hawker et al. defined “severe” arthritis by the 25th percentile of the summary WOMAC score taken from historic data. The cutoff was 39. Thirty-nine falls at the 20th percentile in our
waiting list for arthroplasty, making 80% of our waiting patients eligible Hawker's conservative criterion for severity.

5.3 Limitations

5.3.1 Sample size

There were 147 patients included in the study. The limitation is that the sample size is not sufficiently large to detect an inter-group difference. The power of the chi-square test in our study for the WOMAC function subscale is only 0.26 to detect a 12% difference in the percentage of “better than expected” outcomes between short and long wait groups. However, the results support the hypothesis that prolonged waiting is a significant independent risk factor for the benefits of THA in terms of the WOMAC function subscales. The powers of the chi-square tests in our study for the WOMAC pain and stiffness subscales are both only 0.05 to detect a 3% difference in the percentage of “better than expected” outcomes between short and long wait groups. One is still able to see a trend in the higher proportion of those achieving a “better than expected” outcome in the shorter waiting group. With such a low power, it is hard to achieve any statistical significance from our analysis.
5.3.2 Response Bias

Of the 201 patients who answered our baseline questionnaire, only 147 patients responded to one year follow-up assessment, making the overall response rate 73%. Of the 102 patients in the short wait group (wait <= 6 months), 77 patients responded at follow-up. The response rate was 76% in the short wait group. Of the 99 patients in long wait group (wait >6 months), 70 patients responded at follow-up. The response rate was 71% in the long wait group. Therefore there is a potential response bias in the study. The chi-square test shows that the difference of the response rate between groups is not significant (p=0.444).

According to Kim et al.\textsuperscript{54}, patients who do not respond to follow-up surveys by mail tend to have poorer outcomes than responders. That could explain that why there is a difference in the response rate between long and short wait groups. Future studies need to make an effort to improve the response rates.

5.4 Conclusions

The present study found a strong association between baseline and follow-up WOMAC scores. Using the linear regression method, we provide clinicians and researchers with a simple rule to differentiate “better than expected” and “not better than expected” outcomes for individual patients, given their baseline HRQOL scores.

The main finding of this study is that there is an association between waiting time and the probability for individual patients to achieve a “better than expected” outcome. This trend
is present in all three subscales of WOMAC. The effect on WOMAC function is both statistically and clinically significant. Age, gender, and co-morbidity were also found to be confounders for this effect.

5.5 Future studies

Future directions for research may include:

- Validation of this threshold in new clinical settings.
- Study the effects of waiting time with consideration of cost-effectiveness
- Study health outcomes over a longer period, including subsequent revision surgery and survival of the prosthesis

A randomized clinical trial is currently underway to assess the effect of waiting time on the long-term quality of life outcomes. New studies are being planned and an economic analysis is included in the new studies.
BIBLIOGRAPHY


2 The Arthritis Society web site: (www.arthritis.ca/arthritis/osteoarthritis/)


5 Cabanela ME, Campbell DC 2nd, Henderson ED. Total joint arthroplasty. The hip. Mayo Clin Proc 1979 Sep;54(9):559-63


13 Burton KE, Wright V, Richards J. Patient’s expections in relation to outcome of total hip replacement surgery. ANN Rheum Dis

14 Mahomed, Nizar N.; Barrett, Jane A.; Katz, Jeffrey N.; Phillips, Charlotte B.; Losina, Elena; Lew, Robert A.; Guadagnoli, Edward; Harris, William H.; Poss, Robert; Baron,


16 A. K. Nilsdotter, L.S. Lohmander Age and waiting time as predictors of outcome after total hip replacement for osteoarthritis Rheumatology 2002;41:1261-1267


