2008

# Elderly patients’ ability to recall preoperative health status six weeks following total hip arthroplasty

Jacquelyn Danielle Marsh

# Elderly patients' ability to recall preoperative health status six weeks following total hip arthroplasty.

(Spine title: Elderly patients' ability to recall preoperative health)

(Thesis format: Monograph)

by

Jacquelyn Marsh

Graduate Program in Health & Rehabilitation Sciences
Physical Therapy Field

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

# CERTIFICATE OF EXAMINATION

| | |
|---|---|
| Supervisor | Examiners |
| Dr. Dianne Bryant | Dr. Trevor Birmingham |
| Supervisory Committee | Dr. Bert Chesworth |
| Dr. Steven MacDonald | Dr. Douglas Naudie |

The thesis by

**Jacquelyn Danielle Marsh**

Entitled:

Elderly patients' ability to recall preoperative health status six weeks following total hip arthroplasty

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

Date____________________________

______________________________
Chair of the Thesis Examination Board

## Abstract

We investigated patients' ability to accurately recall their preoperative quality of life following hip replacement surgery.

We randomized consecutive patients aged 55 years or older into two groups. At each assessment patients completed self-report questionnaires (group 1: 4-weeks preoperatively, day-of-surgery, 6-weeks and 3-months postoperatively, group 2: 6-weeks and 3 months postoperatively). At 6 weeks postoperatively, patients completed questionnaires based on their recollection of preoperative health status.

174 patients (mean age 71 years) participated. Agreement between actual and recalled data was excellent for disease-specific questionnaires (ICC range 0.86 to 0.88), and moderate for the generic health measures (ICC range 0.48 to 0.60). The use of recalled ratings had minimal effects on power and sample size.

Patients undergoing total hip arthroplasty can recall their preoperative health status at 6 weeks postoperatively with sufficient accuracy, allowing investigators to improve the efficiency of data collection in this population, with minimal expected loss of statistical power.

*Keywords*: Recall, baseline data collection, elderly population, total hip arthroplasty.

### *Co-Authorship Statement*

With the assistance of Dr. Dianne Bryant and Dr. Steve MacDonald, we designed a randomized controlled trial, in which I was responsible for patient recruitment and conducting all follow-up assessments. I was solely responsible for data entry and data analysis. I wrote the original draft of the manuscript, including interpretation of statistical results (with the assistance of Dr. Bryant) and sent the drafted manuscripts to committee members for their comments and suggestions for the critical revision of the manuscript.

## *Acknowledgement*

I wish to express my sincerest gratitude to Dr. Dianne Bryant, for her supervision, advice, and guidance throughout this entire project. She has provided me with invaluable experiences and opportunities, and her passion and commitment to her work have inspired me to continue in the field of research.

I would like to thank Dr. Steve MacDonald for serving as an important member of my thesis advisory committee, and providing valuable feedback and suggestions throughout the design of the study, and writing of the manuscript.

I would also like to thank the staff at the London Health Sciences Centre who assisted in the conduction of this study, including the participating surgeons for allowing me access to their patients, and the orthopaedic research team for their cooperation, guidance, and assistance.

Finally, I would like to thank my family for their continued support and encouragement, and especially Andy for his constant patience, love, and encouragement throughout.

## *Table of Contents*

*List of Tables*

*List of Figures*

## *List of Appendices*

## Chapter 1: Introduction

Arthritis is one of Canada's most common chronic conditions and is a leading cause of pain, physical disability, reduced quality of life, and use of health care services[1]. Approximately 10% of the adult population suffers from osteoarthritis (OA)[1]. There were approximately 28,045 hip replacements performed in Canada in 2006, representing a 10 year increase of 62%, and a 12% increase compared to 2005[2]. The majority of hip replacements were among patients aged 65 years and older, with the mean age being 67.5 years[2]. Joint replacement surgery is a highly cost-effective procedure which improves the quality of life and functional status of patients suffering from chronic osteoarthritis[1].

Patient self-ratings of quality of life, general health, and functional status are considered one of the preferred methods of evaluating the effects of orthopaedic surgical interventions. In conducting clinical trials, researchers often measure a patient's baseline health status to demonstrate similarities between two groups prior to surgery and to control for any differences at baseline in the analysis of outcome data, increasing the power to demonstrate important between-group differences.

The process of baseline data collection however, can be difficult (usually means an additional visit for patients or coordination with preadmission staff), costly and time consuming, and frequently a substantial proportion of patients who appear to meet the study eligibility criteria prior to surgery prove to be ineligible following surgical examination. If patients can accurately recall their preoperative quality of life following surgery, than it seems reasonable to substitute recalled ratings of baseline health status for prospectively collected baseline ratings, which would contribute to a more efficient use of research staff resources, and greatly decrease patient burden, as only those patients found eligible for study participation would complete baseline assessments.

In a previous study by Bryant et al[3], patients undergoing knee arthroscopy with or without ACL reconstruction were able to accurately recall their preoperative health status two weeks after surgery. The mean age of these patients was 40 years; it is not clear whether or not these results can be generalized to an elderly population. It is possible that hip arthroplasty patients differ from knee arthroscopy patients in their ability to recall. Further, recall assessments done at six weeks following surgery (the usual time that hip arthroplasty patients return to their surgeon for their first postoperative visit), may be less accurate than recall assessments collected two weeks postoperatively.

The purpose of the present study was to determine whether patients aged 55 years or older, undergoing total hip arthroplasty could accurately recall their preoperative quality of life, general health, and functional status at their first postoperative visit.

# Chapter 2: Literature Review

## *2.1 Osteoarthritis of the hip*

Osteoarthritis (OA) is currently the most common type of arthritis, affecting approximately 10% of Canadian adults[1]. OA results from the deterioration of cartilage in the joint resulting in pain, stiffness and instability (Figure 1). The patient experiences progressively worsening pain and loss of range of motion. Non-operative treatments may include anti-inflammatory pain medications, nutritional supplements, activity modification, weight loss, physical therapy, or walking aids. If these methods fail to provide relief for the patient, then surgery may be considered. The decision to perform a total hip arthroplasty depends on the diagnosis and severity of arthritis, as well as several characteristics of the patient including; age, activity level, occupation, medical health, and expectations[4].

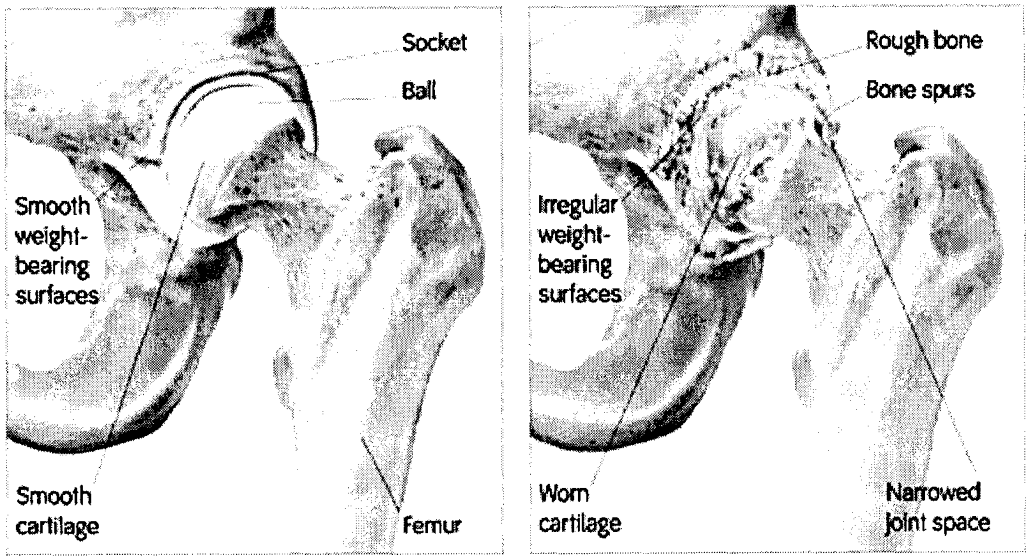


**Figure 1. Comparison of a healthy hip (left) to an osteoarthritic hip (right)[5]**

### *2.2 Total Hip Arthroplasty*

Joint replacement surgery is a highly cost-effective procedure for the treatment of advanced osteoarthritis[1]. A total hip arthroplasty replaces the diseased and painful joint surfaces of the hip with metal and plastic components called a prosthesis. The prosthesis consists of a socket with a plastic liner (replaces natural hip socket), ball (replaces the head of the femur), and stem (inserted into the bone for stability). The procedure is performed through an incision over the side of the hip. First, the end of the femur is cut and replaced with the new ball and stem component. Then, the new socket is inserted in place of the damaged surface and finally the metal ball and socket are joined, allowing for smooth, nearly frictionless motion[6].

The prosthetic hip joint is not as strong or durable as a natural, healthy joint. Synovial fluid helps to lubricate the implants similar to the lubrication of bones and cartilage in a natural hip; however the prosthetic components can still wear or loosen over time. All patients who have had a total hip replacement may need to have a revision at some point, although the length of time that the prosthesis will last varies among patients. Many factors, such as the patient's physical condition, activity level, and body weight also play a role.

A revision hip replacement is a much more complex and costly procedure. Difficulties may arise with the removal of the old prosthesis, as well as with the quality of the existing bone in order to secure the new prosthesis into place. Other complications can also occur as patients tend to be much older when they require a revision and are therefore less tolerant of extensive surgical procedures[7].

Following a total hip arthroplasty, patient self-ratings of quality of life, general health, and functional status are often considered the preferred method of measuring treatment effects. Clinical researchers often collect preoperative baseline measurements to demonstrate similarities between study groups before treatment begins and also to allow for a more powerful statistical comparison of the treatment effect between groups, by adjusting for any differences that were present before treatment began. If the investigators do not have prospective preoperative data, such as in unplanned retrospective study designs, they may rely instead on the patient's recalled rating of their preoperative status. Based on the available literature, it is not clear whether quality-of-life data collected retrospectively accurately represents a patient's actual baseline status collected preoperatively.

### *2.3 Recall and Retrospective Data Collection*

There have been many studies investigating the accuracy of retrospective recalled ratings of health status, however the results are conflicting. Some researchers claim that patients can accurately recall[3,8-14] whereas other studies claim that recalled ratings are highly inaccurate and should not be substituted for prospective health assessments[15-26]. There are two conflicting theories that address the validity of retrospective assessments of subjective outcome measures, the response shift theory[27,28] and the implicit theory of change[29].

### *2.4 Response Shift Theory*

The response shift theory is described as changes in an individual's health status that may produce behavioural, cognitive and affective changes, which have the potential to alter an individual's standards, values or conceptualization of health related quality of life (HRQL). This response shift will consequently influence the patient's perceived quality of life[27, 28].

Response shift theorists argue that patients may change their internal standard of measurement (scale recalibration) as a result of their changing health status. Further, patients' values may change over time so that aspects of quality of life, such as physical health, social health, and psychological health for example, change in the order of their importance. Finally, patients may redefine the target construct (e.g. quality of life) so that how they define that construct changes (scale reconceptualization)[27]. This theory has been used to explain why a patient suffering from a chronic illness may maintain a relatively consistent self-perceived quality of life despite obvious changes in objective or clinical measures of health. Simply put, when a patient is asked to recall a previous health state, they are using their current perceptions about health and it's meaning to rate their previous health state. If their perceptions of health have changed, then it may not be that their recall is inaccurate but rather that the metric from which they are referencing has changed[27].

The response shift theory assumes that the same internal standard of measurement is being used by the patient when making both the recalled assessment and the current health rating. As both measurements are collected at the same point in time, the retrospective assessment is assumed to be a more valid measurement of a prior health state than the actual pretreatment rating, since the patient's perception of health and the metric of measurement has changed since the initial prospective rating[30]. Response shift theorists claim that differences between the actual and recalled preoperative ratings represent evidence of a response shift[27].

### *2.5 Implicit Theory*

The implicit theory of memory, initially described by Michael Ross[29], states that people are aware of the stability of their health status as well as any conditions which might produce a change in their health, such as an intervention or surgery. The implicit theory argues that in order to recall a previous health state the patient must first consider their present state and work backwards, inferring what their initial state must have been[29].

Implicit theorists believe that recalling a previous state is difficult without other contextual features to associate with the memory[29,31]. Without such a reference point people begin their recollection by asking themselves how they are currently, followed by asking themselves how they think things have changed, and then infer what their initial state must have been like[29]. The difference between an actual and a recalled preoperative rating is thought to be a reflection of the error in this process[29].

For example, one of the most common findings in the literature suggests that patients whose health status has improved, or is perceived to have improved following an intervention, recall their prior status as worse than they had previously rated[9,15,19,20,23-26,32]. It is thought that the patient uses their current health status as a reference point, and if they feel that the intervention was effective, they consider their current health and rate their pretreatment health status as worse than their current state[29].

If the patient believes that no change has occurred they may assume that their current health status is the same as their previous health state. This may produce accurate recall in instances where in fact there has been no change. If however, gradual change has occurred so that the patient is unaware of the change, inaccurate ratings are expected. For example, these patients may be inclined to overestimate their prior health status, whereas those who have gradually gotten worse may rate their prior status as lower than it actually was[29].

This trend has also been described by some authors using the cognitive dissonance theory, which suggests that patients who have received an intervention are likely to overstate the benefits of that treatment, especially if the event was stressful, such as surgery[33]. For example, in a study by Mancuso et al.,[25] 104 osteoarthritic patients completed the Hip Rating Questionnaire before undergoing total hip replacement, and were then asked to provide recalled ratings of preoperative status using the same questionnaire at a mean of 2.5 years following surgery. There was poor to fair agreement between prospective and retrospective ratings, with 50% of patients recalling more pain, and 31% recalling worse function than was reported preoperatively.

Although both the response shift theory and the implicit theory consider a patient's ability to recall a previous health state, they each arrive at a different conclusion. Response shift theorists consider the retrospective judgment of a health state as the more valid assessment as the patient derives this judgment from additional information which was not available at the time of the initial prospective evaluation. The implicit theorist argues that the prospective assessment is more valid, as the retrospective test is based on a comparison of present status to an inferred previous health state, which cannot be externally validated[30].

### *2.6 Other factors that affect recall*

There are several other factors that may affect patients' ability to accurately recall preoperative quality of life, general health and functional status postoperatively including the patient's health at the time of the recall task, the nature of the patient's condition (acute versus chronic), the type of outcome being measured, the length of time between actual and recall ratings, the patient's age and surgical effects.

#### *2.6.1 Health status at the time of recall*

The patient's mental health at the time of recall has been shown to predict the accuracy of retrospective assessments. Specifically, patients in poor mental health show a decreased ability to recall compared to patients in good mental health[21, 24, 34-36]. Similarly depression tends to impair recall ability, and has been described by the mood congruency effect, which suggests that the depressed patient is biased towards recalling events that are consistent with the depressed mood[31]. Kent[35] also examined the role of mood and emotional state in the memory of pain among dental patients. The study compared ability to recall pain between two groups of patients, a high anxiety and a low anxiety group. He found that the patients in the high anxiety group overestimated their recalled level of pain, whereas the low anxiety group had a much higher agreement between actual and recalled pain ratings (r=0.79).

Similarly, it has also been suggested that pain state at the time of recall affects the patient's retrospective assessment of pain[21,31,37]. Eich et al.[37] investigated the influence of current pain state on the accuracy of recall of pain intensity. Among patients with chronic headaches who recorded their pain levels in a diary, it was found that the recalled ratings were directly related to the patient's present pain intensity. Recalled assessments made at higher pain levels were significantly greater than those actually obtained at baseline, and the reverse was observed when ratings made during the lowest pain level were examined, with lower actual pain levels being recorded at baseline[37].

### *2.6.2 Acute versus Chronic Health States*

Another common finding among studies is that patients with acute health states can accurately recall prior pain and function[8,9,12,34]. For example, Singer found that recalled assessments of acute painful events one and seven days later were similar and highly correlated with initial assessments using two verbal numeric scales[8]. On the other hand, patients who suffer with chronic health conditions have been shown to provide inaccurate recalled ratings of prior states, tending to overestimate previous pain[15,16,19,24,25,37,38]. For example, Lingard et al.[24] collected recall data for 770 patients, three months after total knee arthroplasty. They found only poor to moderate agreement between actual and recalled ratings, with 31% of patients recalling significantly more pain than they had reported preoperatively[24].

It may be that memory for an acute situation differs from memory for more chronic conditions due to the sudden and unexpected nature of the event, making it distinct and easier to recall[31]. Similarly, patients who have undergone an intervention, such as surgery or therapy, may be able to more accurately recall pre-treatment health, as the intervention itself may serve as a sufficiently significant event to provide a reference point from which to draw accurate recollections of health before the intervention[3]. Patients suffering from chronic conditions may have more difficulty recalling a previous state from several weeks or months earlier if there is no significant event such as an intervention to associate with[30].

### *2.6.3 General versus Specific Health States*

It is also thought that accuracy of recall is affected by the health construct that is being measured[8,18,22,23,25,32]. Specifically, a patient may be able to remember concrete dimensions such as pain more accurately than abstract constructs such as general health as measured by an SF-36 for example[23]. In a study of 200 patients with rheumatoid arthritis, retrospective assessments of pain and global health were collected 2 weeks following treatment with a local corticosteroid injection[22]. It was found that recall bias was larger for the global health assessments, and that assessments of actual and retrospective global health were less strongly correlated than recalled and actual measures of pain[22].

### *2.6.4 Length of Time between Assessments*

Another important factor in the accuracy of recalled ratings may be the amount of time passed between the actual and the retrospective assessments. Studies in which the time interval between actual and recalled reports is relatively short (2 weeks or less) demonstrate that patients can provide an accurate recall of prior health states[3,8,9,11,12,14,22].

For example, Bryant et al.[3] found that patients undergoing knee surgery provided accurate ratings of preoperative quality of life and function 2 weeks after surgery.

On the other hand, studies in which the time interval between actual and recalled ratings is longer (greater than 3 months) show that patients cannot provide an accurate rating of preoperative status through recall[15-17,20,23-26,38]. This finding was demonstrated in the study by Mancuso et al.[25] that tested patients recall ability of preoperative status following total hip arthroplasty. At an average of 2.5 years after surgery patients were found to have poor recall of pain, function, and impact on health. They tended to recall more pain and worse function than what was rated by them prior to surgery.

One possible explanation for these findings was given by Bryant et al.[3] who suggested that during the immediate post-operative period patients may attribute any current pain and disability to the effects of surgery, assuming that their current health status is temporary. In this setting, patients would not use their current state as a reference point from which to judge their previous state, as suggested by the implicit theory of memory, they would instead need to draw from other sources to arrive at a rating of their previous health state. It is possible that patients who have not yet reached their expected rehabilitated state following an intervention may be able to recall their prior health more accurately than if pain and discomfort associated with surgery has already subsided[3].

### *2.6.5 Age*

Some studies have shown that older patients have less agreement between actual and recalled ratings compared to younger patients[16,24,34]. In the study by Lingard et al.[24] for example, patients with a mean age of 70 years (range 38 – 90 years) were asked to recall their preoperative status following total knee arthroplasty. Results showed that those patients aged 75 years or older had significantly poorer recall of both pain and function, compared to the younger patients.

### *2.6.6 Surgical Effects on Recall*

Some authors have suggested that aspects of the surgery itself may have cognitive effects on the patient, such as memory loss, which may hinder a patient's ability to recall preoperative health status. This includes factors such as the type of anaesthetic used (general vs. spinal), the length of time of the surgical procedure, and total blood loss during surgery.

The negative cognitive effects of surgery have been referred to as post-operative cognitive dysfunction (POCD) which is defined as problems with memory, learning and the ability to concentrate after surgery[39]. This condition has been extensively studied among elderly patients, but until recently the majority of studies focused on cardiac surgery patients. Previous studies have suggested that symptoms may persist in some patients for months or years following surgery[40], however the prevalence, causes, risk factors, and consequences of long-term postoperative cognitive dysfunction after non-cardiac surgery are unknown.

The International Study of Post-Operative Cognitive Dysfunction (ISPOCD1)[39] was the first large multicentre study to investigate the occurrence of long-term postoperative cognitive dysfunction in elderly patients after major non-cardiac surgery. The study involved 1,218 patients aged at least 60 years undergoing abdominal or orthopaedic surgery. Patients completed neuropsychological tests before surgery and 1 week and 3 months after surgery. POCD was present in 266 patients (25.8%) 1 week after surgery and in 94 patients (9.9%) 3 months after surgery. It was concluded that the risk of long-term postoperative cognitive decline in the elderly increases with age. Duration of anaesthesia, education level, subsequent operations, postoperative infections, and respiratory complications were also risk factors for early, short term postoperative cognitive dysfunction, but only age was a risk factor for long term postoperative cognitive dysfunction[39].

Several authors have also studied the effect of anaesthetic type on cognitive functioning following surgery[41-44]. Maurer et al.[44] found that patients who received a general anaesthetic had more complications than those who received spinal anaesthesia. The study compared blood loss, operative time, and rate of complications in 606 patients undergoing primary total hip arthroplasty with either spinal or general anesthesia. Patients were followed for 2 years after surgery. Compared with patients receiving a general anaesthetic, the patients who received spinal anaesthesia showed mean reductions in operative time, blood loss, intraoperative transfusion requirements, and had higher hemoglobin levels, all factors which have been shown to reduce the risk of post operative cognitive dysfunction[39].

Other studies however, have found no significant differences in the incidence of long term cognitive dysfunction when comparing general versus regional anaesthesia in elderly patients[41-43]. For example, Rasmussen et al.[42] randomly allocated patients aged 60 years and older undergoing major non-cardiac surgery to receive either general or regional anaesthesia. Cognitive function was assessed preoperatively, and at 1 week and 3 months postoperatively. Although the incidence of POCD at 1 week postoperative was significantly greater in patients who had received general anaesthesia, at 3 months after surgery no significant difference was found in the incidence of cognitive dysfunction between the two groups[42].

## Chapter 3: Study Objectives

Our first objective was to determine the reliability and the validity of recalled ratings. We assumed that, if valid, recalled ratings collected six-weeks postoperatively would accurately predict ratings provided on the day of surgery. Second, we assumed that both the day of surgery and six-week recalled assessments were measuring the same construct and would therefore have high agreement or reliability. Finally, we wished to determine the amount of error in the recall ratings, including both the total variance (between and within-subject variability and random error), as well as individual measurement error.

In anticipation of criticism that might suggest that agreement statistics may be falsely inflated because patients in group 1 have two exposures to the questionnaires prior to being asked to recall, our second objective was to determine whether there was a significant difference in the mean scores and variances of data collected six weeks postoperatively between Group 1 and Group 2 participants. We assumed that because patients were randomized to groups, their average ratings should be similar.

Generally in a test re-test situation, patients with a stable disease are asked to complete self-assessments on two separate occasions, and differences in scores between the two occasions is usually attributed to random error, but may actually consist of both random error and any true change in health status. Since one of the assessments in our study took place on the day of surgery, we hypothesized that a third source of error might arise from anxiety or nervousness on the day of surgery. If accurate recall is possible then random error will be the only source of error suggesting that agreement between recalled and actual ratings should be higher than that observed in a test re-test situation. Therefore, our third objective was to calculate the agreement between ratings from the day of surgery and 4 week preoperative data and compare them to the agreement between the day of surgery and recall data (Group 1).

Finally, for many investigators, the purpose of collecting baseline data is to use the data as a covariate when testing for significant differences between groups. In order for a covariate to contribute to a reduction of the unknown variance, it must have a correlation to the outcome of interest greater than 0. Therefore, the next set of objectives was to determine the correlation between prospectively collected baseline data and 3 month post-operative data. If the correlation is not greater than 0, then collection of baseline data is not necessary for this population. If it is greater than 0, our final objective was to compare the correlation between actual baseline and postoperative data to the correlation between retrospective baseline data and postoperative data and to assess the efficiency of using retrospective data by analyzing its effect on sample size and power.

## Chapter 4: Methods

### *4.1 Study Design*

This prospective, randomized clinical trial was conducted at London Health Sciences Centre with five orthopedic surgeons participating in patient recruitment. Patients who were scheduled for a total hip arthroplasty were contacted at least 4 weeks prior to surgery to determine their willingness to participate in the study. Consenting patients were randomly allocated into one of two groups. Group 1 underwent assessment at four weeks preoperatively, on the day of surgery, and at six weeks and three months postoperatively. Participants allocated to Group 2 underwent assessment at six weeks and at three months postoperatively. At each assessment patients were asked to complete several self-report questionnaires including disease-specific quality of life, general health, and functional status (Figure 2).

At six weeks postoperative, patients in both groups were provided with two sets of questionnaires. For the first set of questionnaires, the patient was asked to recall their quality of life, general health, and function during the period four weeks prior to surgery and to complete the questionnaires according to that recall. After completion of the recalled version, the patient was then asked to assess their current quality of life, general health, and functional status over the past 2 weeks. Six weeks was selected because it reflects the usual timeframe that a patient returns for their first postoperative visit and therefore represents a time that would be the least burdensome for patients and research staff to complete baseline assessments should recall be shown to be sufficiently accurate.

Finally, at three months postoperative, each patient completed the questionnaires to assess current quality of life and health status during the past 2 weeks (Figure 2).

In anticipation of the possibility that patients who completed the questionnaires on previous visits would actually remember their previous responses (producing agreement statistics between actual and recalled ratings that were falsely inflated), we randomly assigned patients into one of two groups; one group that would provide ratings before being asked to recall and one group that would not. Since patients were randomly assigned to groups, participants in group 1 and group 2 were assumed to be similar and therefore should have similar recall ratings. There are two possible outcomes that we may observe if experience with the questionnaires influences patients' ability to recall (presumably improving recall ability): 1) a systematic difference between groups for the recall ratings only (i.e. the systematic difference will not be present in the other postoperative ratings), or 2) a greater variability between patients within group 1 for recalled ratings which may be evidenced by heterogeneous variances between groups.

### *4.2 Eligibility Criteria*

Patients included in the study were aged 55 years or older, and undergoing either primary or revision hip replacement surgery for the treatment of osteoarthritis. Participants currently enrolled in clinical trials using similar questionnaires were excluded to decrease any learning effect. We also excluded patients undergoing minor procedures such as removal of hardware, or those with rare diseases or conditions. Patients with no fixed address, those with a major psychiatric illness, those who were cognitively impaired, and those unable to speak or understand English were also excluded.

Randomization was stratified by surgeon and the type of surgery being performed (primary or revision) to balance potential prognostic characteristics between groups. The randomization sequence was constructed using a computer algorithm with permuted block sizes of three and six. To ensure adequate concealment of allocation, the study coordinator established patient eligibility and obtained verbal consent prior to randomization.

### *4.3 Outcome Measures*

Questionnaires included the Lower Extremity Functional Scale (LEFS), the Western Ontario and McMaster Universities Osteoarthritis index (WOMAC), The Oxford Hip Score, the Short-Form Health Survey (SF-12), and the Feeling Thermometer.

The LEFS[45,46] is a 20-item, region-specific quality-of-life questionnaire for patients with lower limb pathologies. Each item has five response options, ranging from 0 to 4, with 0 representing extreme difficulty or inability to perform the activity, and 4 representing no difficulty in performing the activity. Each question has a maximum score of 4, with a higher score indicating a greater functional level. A change in score of at least 9 points is considered a clinically important difference. The LEFS has face validity, and has demonstrated construct validity, reliability, and responsiveness[45,46].

The WOMAC[47-49] is a 24-item, disease-specific questionnaire. The index consists of 24 questions, divided into three domains: pain, stiffness, and difficulty with physical function. Individual questions are assigned a score between 0 points (no pain, stiffness, or difficulty with physical functions) and 4 points (extreme pain, stiffness, or difficulty with physical functions). Domains are equally weighted and reported as sums, with a higher number indicating a greater burden of osteoarthritis. The WOMAC is extensively used and has been shown to be a valid, reliable instrument that is sensitive to change[47-49].

The Oxford Hip Score[50-52] is a 12-item, disease-specific quality-of–life questionnaire for patients undergoing total hip replacements, designed specifically to capture joint arthroplasty outcomes. The total score is summed from the 12 responses to the questions which address joint pain, function, and mobility. Scores for each item range from one to five with a higher score indicating a poorer health state. This instrument has face validity and has demonstrated construct validity, reliability, and is sensitive to change[50-52].

The SF-12[53] is a 12-item generic health instrument that evaluates eight domains including restrictions or limitations on physical and social activities, normal activities and responsibilities of daily living, pain, mental health and well-being, and perceptions of health. The SF-12 has been extensively used, and has been shown to be valid, reliable, and responsive in a wide variety of populations and contexts including patients with orthopedic conditions[53].

The Feeling Thermometer[54-56] is a visual analogue scale presented in the form of a thermometer with 100 intervals, in which the best state is full health (equal to a score of 100) and the worst state is death (a score of 0). This instrument has face validity, and has demonstrated construct validity, reliability, and has also shown responsiveness to change[54-56].

All questionnaires were transformed to a 100 point scale to ease in comparison across scores. At each assessment, the patients were asked to consider their quality of life, health, and functional status during the past four weeks. For the recall assessment, we asked patients to consider their health status during the four weeks prior to their surgery, and respond to the questionnaires based on that recall. Therefore we considered the ratings provided on the day of surgery to be the gold or criterion standard of patients' preoperative health, and thus recalled ratings collected 6 weeks postoperatively should accurately predict ratings provided on the day of surgery. It was also assumed that both time points measure the same construct and would therefore have high agreement or reliability.

### *4.4 Sample Size*

To provide estimates of agreement between the recalled and actual data, the appropriate calculation to determine sample size requirement is one that allows us to estimate a parameter (0.85) with a pre-specified level of precision, with a 95% confidence interval no wider than 0.10. Using sample size calculations for estimating a parameter[57], we needed 111 participants in Group 1. To ensure adequate power (80%) for a between-group comparison of the recalled and current ratings between Group 1 and Group 2 to assess the similarities between groups addressing our second objective, we required approximately 50 patients per group, given a Type 1 error rate of 0.05. Thus, we randomized patients using a 2:1 randomization schedule (Group 1: Group 2).

### *4.5 Statistical Analysis*

In keeping with our first objective, to determine the validity of the recalled ratings, we performed a linear regression to determine the ability of patients' recalled data to predict the day of surgery ratings for each of the questionnaires. We then constructed scatterplots of the data with 95% prediction lines to explore the variability (between- and within-subject) and agreement between the two ratings at the group and individual levels.

To determine the reliability, we conducted a repeated measures analysis of variance (ANOVA) to determine whether there were any significant systematic differences between the day of surgery and the recall ratings. To estimate the magnitude of the association between recalled and actual preoperative data, an intraclass correlation coefficient (ICC) (two-way mixed model with measures of absolute agreement) for each instrument and its 95% confidence interval was constructed using the mean square values of between-subjects (patients), within-subjects (time), and error from the ANOVA. The ICC provides information about the total variance (between and within-subject variability and random error), whereas the standard error of measurement (S.E.M.) expresses individual measurement error only, without the influence of variance among patients[58]. Therefore we also calculated the S.E.M. and its 95% confidence intervals from the ANOVA (square root of the mean square error).

In keeping with our second objective, we compared the recalled ratings between participants in group 1 and group 2 using independent samples t-tests to determine whether the two prior exposures to the questionnaires of participants in group 1 (four week preoperative and day of surgery) had an influence on their recalled ratings. We assumed that because patients were randomized into groups, participants in both groups would be similar with respect to their baseline health status and that if accurate recall was possible, the scores for the recalled data would also be similar between groups. Further, to determine if there were any significant differences between the variances of ratings between groups, we used Levene's Test for equality of variances[59,60] where a significant test ($p<0.05$) indicated unequal variances.

In keeping with our third objective, to determine the test-retest reliability of patient ratings provided at the preadmission appointment and on the day of surgery, we constructed an intraclass correlation coefficient (ICC) (two-way mixed model with measures of consistency) for each instrument, as well as scatterplots with 95% group and individual level prediction intervals.

In keeping with our final objective, to explore the efficiency of using recalled data in place of the prospectively collected baseline data, the effect on power and sample size estimates were determined for each of the questionnaires. This was done according to three common methods of making statistical comparisons between groups: 1) a t-test of the posttest score only, 2) a t-test of the change score (pretest-posttest), and 3) an analysis of covariance (ANCOVA), with actual or recalled pretest scores used as the covariate. For all calculations, the probability of type I error was maintained at 0.05, and the probability of type II error at 0.20. A difference of 20% of the mean preoperative score was considered an important difference.

## Chapter 5: Results

### *5.1 Patient Characteristics*

We assessed the eligibility of 221 consecutive patients who were scheduled for either a primary or revision total hip arthroplasty, of which 39 did not participate. Nineteen patients refused consent, 4 were non-English speaking, 6 patients were having ineligible procedures and 10 cancelled surgery; leaving 182 patients who gave consent and were randomized. One hundred and twenty-one patients were randomized to group 1 (105 primary, 16 revision), and 61 patients were randomized to group 2 (50 primary, 11 revision). Eight patients were excluded following randomization (5 were not eligible for surgery, 1 patient was hospitalized for other medical complications, 1 patient was lost to follow-up, and 1 patient died following surgery). Therefore 174 patients were included in the analysis (group 1=118, group 2=56) (Figure 2).

The mean age of study participants was 70.6 years (range 55 years to 91 years), and the majority of patients (87.8%) were retired. Eighty-three percent of patients had a primary total hip arthroplasty, while 17% had a revision. Patients were similar in age, gender, operative hip, type of procedure being performed (primary or revision hip arthroplasty), and prevalence of previous hip surgery between group 1 and group 2 participants. Table 1 provides a detailed description of the demographic characteristics of the study participants.

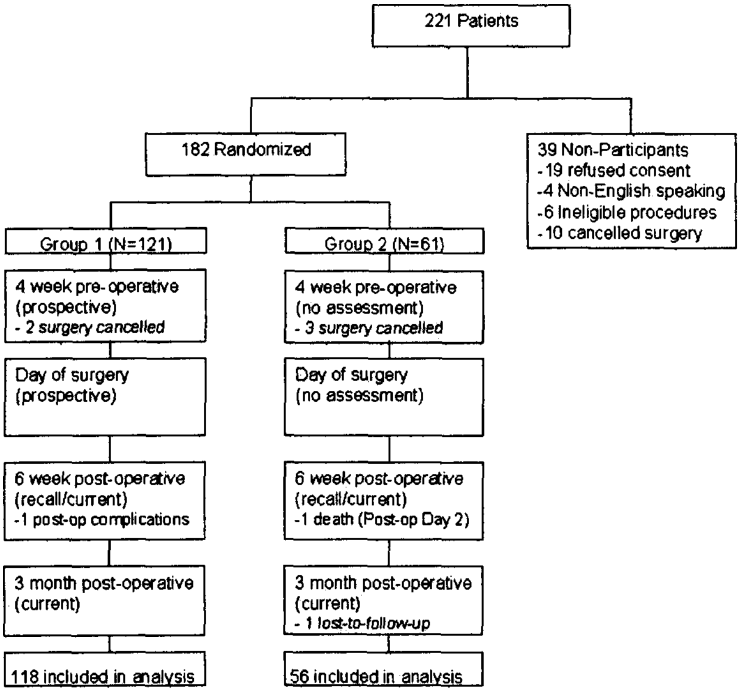


**Figure 2. Flow of participants through the trial**

**Table 1. Demographic characteristics of participants**

| Characteristic | Group 1 (n=118) | Group 2 (n=56) |
|---|---|---|
| Female | 54.7% | 41.8% |
| Age (years) | 70.9 ± 8.9 | 69.9 ± 7.9 |
| Height (inches) | 66.2 ± 4.1 | 67.6 ± 4.2 |
| Weight (pounds) | 182.1 ± 40.0 | 184.5 ± 28.0 |
| Smoking status | | |
| Never smoked | 50.9% | 57.9% |
| Smoked, but quit | 34.2% | 28.9% |
| Current smoker | 14.9% | 13.2% |
| Right hip affected | 53.4% | 50.0% |
| Primary THA | 85.6% | 82.5% |
| Revision THA | 14.4% | 17.5% |
| Anaesthetic Type | | |
| General | 40.9% | 49.1% |
| Spinal | 59.1% | 50.9% |
| Anaesthetic time (min) | 125.6 ± 33.5 | 131.51 ± 43.4 |
| Previous surgery | 27.4% | 24.1% |
| Symptoms in other hip | 43.6% | 43.9% |
| SFmental health domain | 52.6 ± 9.3 | 50.3 ± 8.1 |
| Employment status | | |
| Retired | 87.0% | 90.2% |
| Employed | 10.4% | 9.8% |
| Unemployed | 1.7% | 0.0% |
| Disability | 0.9% | 0.0% |

### *5.2 Objective 1: Patients' ability to recall preoperative quality of life and general health status*

The mean differences between actual baseline ratings collected on the day of surgery and recalled ratings provided at 6 weeks postoperatively were small across all questionnaires. Three of the differences were statistically significant (WOMAC =2.80 (0.73 to 4.88); SF-12 Physical component score = -2.82(-4.39 to -1.27); and Feeling Thermometer =5.17 (1.72 to 8.61)(Table 2).

**Table 2. Agreement between Day of Surgery and Recalled baseline data**

| Questionnaire | Mean Difference (95% C.I.) | p-value | ICC (95%C.I.) | S.E.M. (95% C.I.) |
|---|---|---|---|---|
| LEFS | -0.97(-2.36 to 0.41) | 0.17 | 0.86 (0.79 to 0.90) | 4.70 (4.09 to 5.51) |
| OHS | -0.32 (-1.92 to 1.28) | 0.69 | 0.87 (0.81 to 0.91) | 5.23 (4.57 to 6.14) |
| WOMAC | 2.80 (0.73 to 4.88) | <0.01 | 0.88 (0.82 to 0.92) | 7.11 (6.20 to 8.35) |
| SF-12 (PCS) | -2.82 (-4.39 to -1.27) | <0.01 | 0.58 (0.40 to 0.71) | 4.98 (4.34 to 5.83) |
| SF-12 (MCS) | 2.04 (-0.42 to 4.50) | 0.10 | 0.48 (0.30 to 0.62) | 8.40 (7.33 to 9.85) |
| FT | 5.17 (1.72 to 8.61) | <0.01 | 0.60 (0.43 to 0.72) | 11.44 (9.99 to 13.40) |

*Abbreviations:* LEFS = Lower Extremity Functional Scale, OHS=Oxford Hip Score, WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index, SF-12=Short-Form Health Survey, PCS=Physical Component Score, MCS=Mental Component Score, FT=Feeling Thermometer.

Scatterplots of group 1 patients' recalled versus day of surgery data were suggestive of high levels of agreement (Figure 3). The data were also consistent with the assumptions of linear regression (linearity, normality, and homoscedasticity) as verified by the residual analysis.

Recalled ratings were a significant predictor of actual baseline ratings (p<0.001) across all questionnaires (Table 3). Pearson's correlation coefficient indicated excellent agreement between ratings for the region-specific measures (LEFS r =0.86, Oxford Hip Score r =0.87, and WOMAC r =0.89). The correlation between actual and recalled ratings of the generic health measures was moderate (SF-12 PCS r =0.62; SF-12 MCS, r =0.48; and Feeling Thermometer, r =0.63) (Table 3).

**Table 3. Predictive validity of using retrospective (recalled ratings) in place of prospective (day of surgery) ratings**

| Questionnaire | Pearson's *r* | Coefficient (B) |
|---|---|---|
| LEFS | 0.86 | 0.80 (0.70 to 0.91), p<0.001 |
| Oxford Hip Score | 0.87 | 0.81 (0.71 to 0.90), p<0.001 |
| WOMAC (Total) | 0.89 | 0.84 (0.75 to 0.93), p<0.001 |
| SF-12 (PCS) | 0.62 | 0.54 (0.39 to 0.68), p<0.001 |
| SF-12 (MCS) | 0.48 | 0.50 (0.31 to 0.70), p<0.001 |
| Feeling Thermometer | 0.63 | 0.51 (0.38 to 0.64), p<0.001 |

*Abbreviations:* LEFS = Lower Extremity Functional Scale, WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index, SF-12=Short-Form Health Survey, PCS=Physical Component Score, MCS=Mental Component Score.

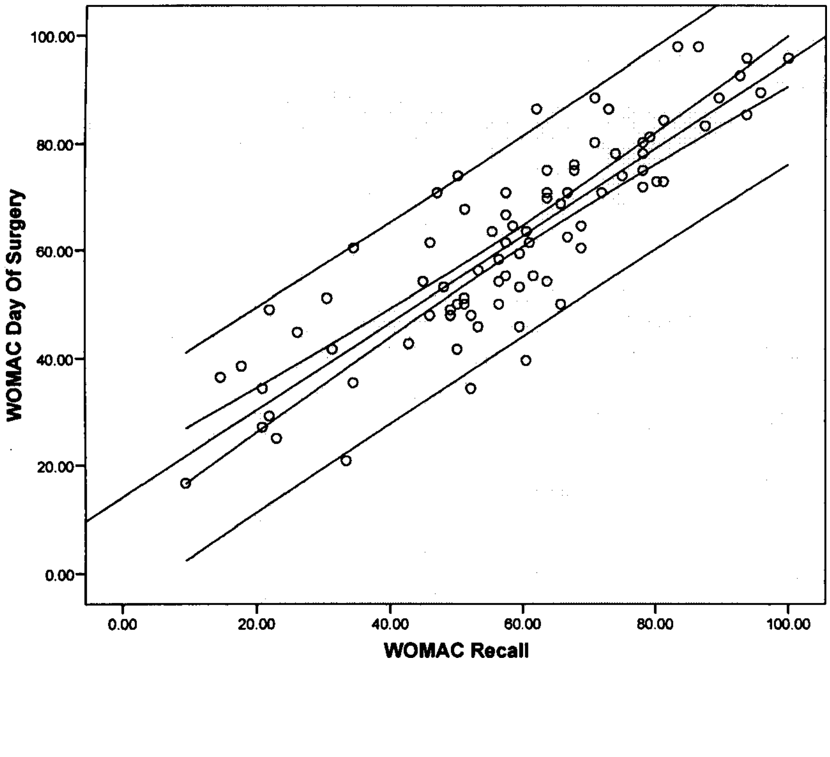
Similarly, the agreement between recalled ratings and day of surgery ratings was excellent across the disease-specific questionnaires (LEFS ICC=0.86, 95% CI 0.79 to 0.90; Oxford Hip Score ICC=0.87, 95% CI 0.81 to 0.91; WOMAC ICC=0.88, 95% CI 0.82 to 0.92), whereas agreement for the generic health questionnaires was moderate (SF-12(PCS) ICC=0.58, 95% CI 0.40 to 0.71; SF-12(MCS) ICC=0.48, 95%CI 0.30 to 0.62; Feeling Thermometer ICC=0.60 95%CI 0.43 to 0.72) (Table 2).

The standard error of measurement was relatively small for both the disease-specific and generic health questionnaires (Table 2), suggesting that the lower levels of agreement between the day of surgery ratings and the six-week postoperative recalled ratings of the generic health measures (SF-12 (PCS) ICC = 0.58, SF-12(MCS) ICC=0.48, Feeling Thermometer ICC=0.60) is due to smaller between-subject variability, or less heterogeneity in scores, rather than to a greater degree of error.

Figure 3 displays scatterplots with 95% mean and individual prediction lines for the WOMAC (an example of large between-subject variability) and the MCS of the SF-12 (an example of small between-subject variability) (Figure 3). The SF-12 MCS scores of patients in our study population fell within the middle part of the scale, indicating that they do not represent the entire range of scores possible for the SF-12 among the general population. The disease or region-specific questionnaires (LEFS, WOMAC, Oxford Hip Score) show a larger between-subjects effect, representing a greater proportion of the possible scores among a hip arthroplasty population, and therefore display greater between-subject variability, as displayed in the WOMAC (Figure 3).

**Figure 3a-b. Scatterplots with 95% Group and Individual Prediction Intervals.**

**a)**



**b)**

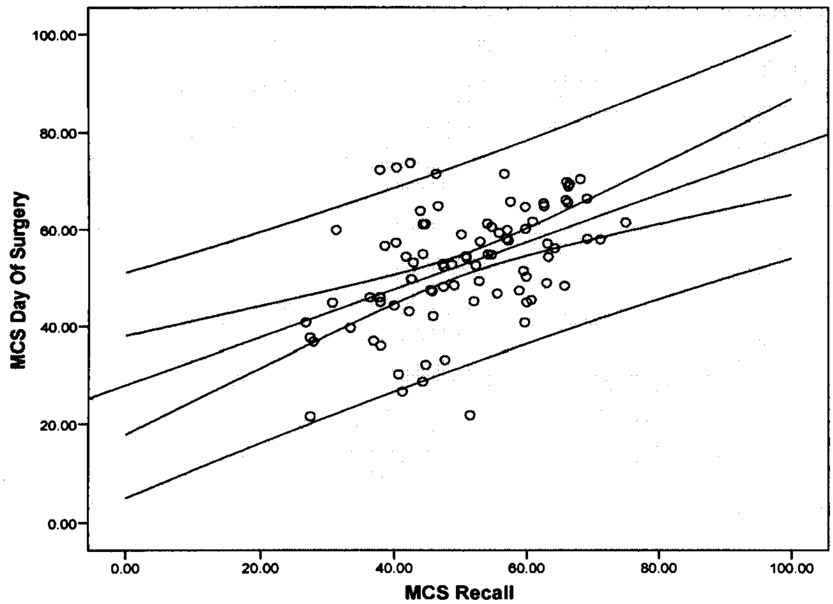


**Fig 3.** Patients' recalled ratings of quality of life compared to actual ratings for the a) Western Ontario and McMaster University Osteoarthritis Index (WOMAC), and b) Short-Form Health Survey (SF-12) Mental Component Score (MCS).

### *5.3 Objective 2: The influence of group 1 participants' prior exposure to instruments on ability to recall*

The independent samples t-test comparison between group 1 and group 2 participants' recalled ratings (6 weeks postoperative) was not significant for any of the questionnaires (Table 4). The mean differences between group 1 and group 2 recalled ratings were small across all instruments, with only the WOMAC total score reaching statistical significance with a mean difference of -7.96 (95%CI -14.52 to -1.40, p=0.02), which could be considered a spurious finding since the difference between groups did not reach significance for any other of the questionnaires. The variances of recalled ratings were also similar between groups across questionnaires with only two differences reaching statistical significance (SF-12 physical component score, p<0.01, and Feeling Thermometer, p=0.05) (Table 5).

**Table 4. Mean scores (± standard deviation) of questionnaires at all time points for both groups**

| Time | Questionnaire | Group I<br>(n=118) | Group II<br>(n=56) |
|---|---|---|---|
| 4-Week Pre-operative | LEFS | 15.7 ± 12.7 | |
| | OHS | 67.2 ± 14.2 | |
| | WOMAC – Total | 54.9 ± 18.2 | |
| | SF-12 PCS | 27.4 ± 8.8 | |
| | SF-12 MCS | 53.1 ± 11.8 | |
| | Feeling Thermometer | 58.7 ± 19.3 | |
| Day of surgery | LEFS | 12.7 ± 11.6 | |
| | OHS | 70.5 ± 14.3 | |
| | WOMAC – Total | 60.3 ± 19.7 | |
| | SF-12 PCS | 25.0 ± 7.8 | |
| | SF-12 MCS | 53.1 ± 11.8 | |
| | Feeling Thermometer | 58.0 ± 16.3 | |
| 6-Week Post-operative (Recalled) | LEFS | 14.7 ± 12.9 | 12.9 ± 14.5 |
| | OHS | 70.0 ± 15.2 | 72.9 ± 14.1 |
| | WOMAC – Total | 56.3 ± 20.2 | 64.2 ± 20.3 |
| | SF-12 PCS | 28.0 ± 8.9 | 26.0 ± 6.3 |
| | SF-12 MCS | 50.9 ± 11.4 | 48.5 ± 10.0 |
| | Feeling Thermometer | 54.5 ± 20.1 | 55.0 ± 17.1 |
| 6-week post-operative (current) | LEFS | 16.6 ± 12.2 | 14.6 ± 12.8 |
| | OHS | 55.4 ± 14.6 | 59.3 ± 12.2 |
| | WOMAC – Total | 44.0 ± 17.4 | 47.8 ± 14.3 |
| | SF-12 PCS | 30.5 ± 8.1 | 27.5 ± 6.8 |
| | SF-12 MCS | 54.2 ± 11.5 | 53.3 ± 9.7 |
| | Feeling Thermometer | 68.4 ± 16.6 | 66.9 ± 12.2 |
| 3-month post-operative | LEFS | 33.9 ± 15.4 | 34.4 ± 15.4 |
| | OHS | 39.8 ± 13.1 | 39.5 ± 11.9 |
| | WOMAC – Total | 27.5 ± 15.6 | 26.6 ± 11.6 |
| | SF-12 PCS | 38.5 ± 9.7 | 36.7 ± 9.1 |
| | SF-12 MCS | 56.1 ± 7.8 | 54.0 ± 10.5 |
| | Feeling Thermometer | 75.4 ± 14.3 | 73.3 ± 12.6 |

*Abbreviations:* LEFS = Lower Extremity Functional Scale, OHS=Oxford Hip Score, WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index, SF-12=Short-Form Health Survey, PCS=Physical Component Score, MCS=Mental Component Score.

**Table 5. Assessment of the Similarity between Group 1 and Group 2**

| | Recalled Ratings Mean (SD) | Mean Difference between Groups at Recall (95% C.I.) | P-Value | Levene's Test for equality of variances |
|---|---|---|---|---|
| **LEFS** | | 1.77 (-2.59 to 6.13) | 0.40 | 0.35 |
| Group 1 | 14.7 (12.9) | | | |
| Group 2 | 12.9 (14.5) | | | |
| **OHS** | | -2.95 (-7.74 to 1.83) | 0.23 | 0.42 |
| Group 1 | 70.0 (15.2) | | | |
| Group 2 | 72.9 (14.1) | | | |
| **WOMAC** | | -7.96 (-14.52 to -1.40) | 0.02 | 0.45 |
| Group 1 | 56.3 (20.2) | | | |
| Group 2 | 64.2 (20.3) | | | |
| **SF-12 (PCS)** | | 1.98 (-0.35 to 4.32) | 0.14* | 0.01 |
| Group 1 | 28.0 (8.9) | | | |
| Group 2 | 26.0 (6.3) | | | |
| **SF-12 (MCS)** | | 2.34 (-1.17 to 5.85) | 0.19 | 0.08 |
| Group 1 | 50.9 (11.4) | | | |
| Group 2 | 48.5 (10.0) | | | |
| **FT** | | -1.21 (-6.84 to 5.00) | 0.85* | 0.05 |
| Group 1 | 54.4 (20.1) | | | |
| Group 2 | 55.4 (17.1) | | | |

*Abbreviations:* LEFS = Lower Extremity Functional Scale, OHS = Oxford Hip Score, WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index, SF-12=Short-Form Health Survey, PCS=Physical Component Score, MCS=Mental Component Score, FT=Feeling Thermometer.

*Because these data did not meet the assumption of equality of variances ($p<0.05$), we used the Brown-Forsythe and Welch statistic, which are calculations of ANOVA significance by adjusting results for unequal variances. The p-values obtained from these tests were not different from those obtained by the ANOVA, therefore we reported the p-values from the ANOVA.

### *5.4 Objective 3: Test-Retest Reliability*

Reliability between ratings provided at the 4 week preoperative assessment and on the day of surgery were excellent across all questionnaires (LEFS=0.97, Oxford Hip Score=0.93, WOMAC=0.96, SF-12 PCS=0.82, MCS=0.91, FT=0.94) (Table 6).

**Table 6. Test-Retest Reliability between ratings provided 4 weeks preoperatively and on the day of surgery.**

| Questionnaire | ICC | 95% CI |
|---|---|---|
| LEFS | 0.97 | 0.95 to 0.98 |
| Oxford Hip Score | 0.93 | 0.89 to 0.95 |
| WOMAC | 0.96 | 0.95 to 0.98 |
| SF-12 (PCS) | 0.82 | 0.73 to 0.88 |
| SF-12 (MCS) | 0.91 | 0.87 to 0.94 |
| FT | 0.94 | 0.90 to 0.96 |

*Abbreviations:* LEFS = Lower Extremity Functional Scale, WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index, SF-12=Short-Form Health Survey, PCS=Physical Component Score, MCS=Mental Component Score, FT=Feeling Thermometer.

### *5.5 Objective 4: Effect on sample size and power when using recall data*

For each questionnaire, the correlation between the day of surgery ratings and the 3 month postoperative score did not differ significantly from the correlation between the recalled rating and 3 month postoperative score. The correlation between actual preoperative ratings (day of surgery), and 3 month postoperative ratings ranged from 0.42 to 0.54, whereas the correlation between recalled preoperative and 3 month postoperative ratings ranged from 0.39 to 0.58 across questionnaires (Table 7).

The impact of using recalled ratings in place of the prospectively collected baseline data would result in an increase in sample size ranging from 0 to 41% to detect an important difference if the planned statistical comparisons involve the use of a change score. The required increase in sample size when using an ANCOVA for statistical comparison varied between 3% to 45% if using recalled data in place of actual baseline ratings (Table 7).

All of the sample size estimates using recalled or actual data for a planned ANCOVA were smaller than those that would be required for comparisons using a post-test only score, whereas all of the calculations for sample size using recalled data for comparisons using change scores were greater than those using a post-test only score (Table 7).

Similarly, substituting recalled ratings for prospectively collected baseline data has an impact on power. If using change scores the reduction in power estimates ranged between 0 to 14%, and power reductions from 1% to 13% were estimated if using recalled ratings in place of actual baseline ratings for ANCOVA statistical comparisons (Table 7).

When using an ANCOVA statistical comparison, all estimates of power were greater than the 80% power of a planned post-test only comparison (increase in power between 7% to 11% if using prospective baseline data, or 1% to 9% if using recalled ratings), and also greater than the change score power estimates, with an increase in power between 8% to 13% if using actual data, or between 10% to 15% increase if using recall data (Table 7).

**Table 7. Assessment of the effect of using recalled ratings on sample size and power for three common methods of making statistical comparisons.**

For sample size estimations, we held power constant at 80%. For power estimations, sample size was held constant, using the sample size equation for a post-only score.

| | | | | | Sample size estimations actual:recalled (change in sample size) | | |
|---|---|---|---|---|---|---|---|
| **Questionnaire** | **SDa** | **SDr** | **r1** | **r2** | **Post** | **Change** | **ANCOVA** |
| LEFS | 11.6 | 12.9 | 0.44 | 0.47 | 327 | 367:429 (17%) | 264:315(19%) |
| OHS | 14.3 | 15.2 | 0.48 | 0.39 | 16 | 17:22 (33%) | 12:15(25%) |
| WOMAC | 19.7 | 20.2 | 0.42 | 0.44 | 42 | 48:50 (3%) | 34:36 (4%) |
| SF-12 (PCS) | 7.8 | 8.9 | 0.42 | 0.53 | 38 | 45:46 (4%) | 32:36(12%) |
| SF-12 (MCS) | 11.8 | 11.4 | 0.45 | 0.41 | 19 | 21:21 (0%) | 16:15 (3%) |
| Feeling Thermometer | 16.3 | 20.4 | 0.54 | 0.58 | 42 | 29:40 (41%) | 22:32(45%) |
| | | | | | **Power estimations actual:recalled (change in power)** | | |
| LEFS | 11.6 | 12.9 | 0.44 | 0.47 | 0.80 | 0.75:0.69 (-7%) | 0.88:0.81(-6%) |
| OHS | 14.3 | 15.2 | 0.48 | 0.39 | 0.80 | 0.78:0.67(-12%) | 0.89:0.82(-7%) |
| WOMAC | 19.7 | 20.2 | 0.42 | 0.44 | 0.80 | 0.74:0.73(-1%) | 0.87:0.86(-1%) |
| SF-12 (PCS) | 7.8 | 8.9 | 0.42 | 0.53 | 0.80 | 0.74:0.72(-2%) | 0.87:0.83(-4%) |
| SF-12 (MCS) | 11.8 | 11.4 | 0.45 | 0.41 | 0.80 | 0.76:0.76(-0%) | 0.88:0.89(-1%) |
| FT | 16.3 | 20.4 | 0.54 | 0.58 | 0.80 | 0.83:0.69(-14%) | 0.91:0.79(-13%) |

*Abbreviations:* LEFS = Lower Extremity Functional Scale, OHS=Oxford Hip Score, WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index, SF-12=Short-Form Health Survey, PCS=Physical Component Score, MCS=Mental Component Score, FT=Feeling Thermometer.

*Note*: SDa= standard deviation of actual (day of surgery) baseline data, SDr= standard deviation of recalled baseline data, r1 = Pearson's correlation coefficient of actual baseline ratings (day of surgery) to posttest (3 month postoperative), r2= Pearson's correlation coefficient of recalled baseline ratings (6-week recall) to posttest (3-month postoperative). For all calculations, probability of Type I error = 0.05, probability of type II error = 0.20. An important difference was calculated as 20% of the mean preoperative (day of surgery) rating.

## Chapter 6: Discussion

Patient self-ratings of quality of life, general health, and functional status are considered one of the preferred methods of assessing outcome following total hip arthroplasty. Researchers often measure a patient's baseline health status to demonstrate similarities between two groups prior to surgery and to allow for a more powerful statistical comparison of treatment effect. Baseline data collection however, can potentially introduce large inefficiencies to data collection (unnecessary use of research staff resources, and patient burden), if patients who provide baseline measurements are found to be ineligible for the study following surgical examination. Our study shows that elderly patients undergoing total hip arthroplasty can accurately recall their preoperative quality of life, general health, and functional status 6 weeks postoperatively, providing the potential for a more efficient method of baseline data collection.

Previous studies investigating the accuracy of retrospective data collection have had mixed results, with some investigators supporting the use of recall data[3,8-14] whereas others do not[15-23,25-27]. One factor that may contribute to these discrepancies is the amount of time between the prospective and retrospective assessments. For example, researchers who asked patients to provide recalled ratings less than 2 weeks after an intervention found that patients can accurately recall their preoperative health status[3,8,9,11,12,14,22], whereas those studies which use recalled patient ratings 2 months or longer following an intervention did not find high agreement between the prospective and retrospective ratings[15-17,20,23-26,38].

Based on the implicit theory of memory, we could hypothesize that patients who have not yet reached their expected rehabilitated state following surgery may be able to recall their prior health more accurately than if pain and discomfort associated with surgery has already subsided[3]. Our study asked patients to provide recall ratings at their first postoperative appointment (about six weeks following surgery), suggesting that at this time patients may attribute any current pain and disability to the effects of surgery, expecting their current health status to change, which may force them to use means other than their current state of health to recall their previous health status[29].

The opposing theories put forth by response shift theorists[27, 28] and implicit theorists[29] are useful when analyzing potential reasons for conflicting results in the literature concerning patients' ability to recall. Response shift theorists argue that patients may change their internal standard of measurement (scale recalibration) and the importance of their values concerning health as a result of their changing status, which may also cause them to redefine their quality of life (scale reconceptualization)[27]. Therefore when a patient is asked to recall a previous health state, they use their current perceptions about health and it's meaning to rate their previous health state. If their perceptions of health have changed, then it may not be that their recall is inaccurate but rather that the metric from which they are referencing has changed[27]. In our study, it is possible that at the time of recall patients attributed their current pain and disability to the effects of surgery, assuming it to be a temporary state, and therefore did not use their current health as a reference point from which to judge their preoperative health status[3], suggesting that a response shift had not occurred.

On the other hand, implicit theorists believe that recalling a previous state is difficult without any other events to associate with the memory. Without such a reference point people begin their recollection by asking themselves how they are currently, followed by asking themselves how they think things have changed, and then infer what their initial state must have been like[29]. Since our study patients were still in the rehabilitative state at the time of recall, they were not using their current health status as a reference point to recall their preoperative health, as this was assumed to be temporary. It is possible that the surgery served as a sufficiently significant event, and therefore acted as a reference point for patients to assess how they were currently and then infer what their preoperative health status must have been before their operation.

Although our results support the use of retrospective baseline data collection in this population, it would seem that recall data is most appropriate for group comparisons, as in a clinical trial, in which data are aggregated and then generalized to the population. If however, the clinician is interested in using a patient's preoperative health status to predict their outcome following surgery; our study demonstrates greater uncertainty in the ability to predict outcomes at the individual level when using recalled ratings. This finding is illustrated in our scatterplots, which presents both the group and individual prediction lines (Figure 3a-b).

It is also important to consider the observed versus expected relationship between actual and recalled ratings. If we consider a test re-test situation in which patients with a stable disease are asked to complete a self-assessment on two separate occasions, differences in scores are attributed to random error and, to a much lesser degree, error due to true change. It is also possible that because one of the assessments took place on the day of surgery, there may have been additional error due to anxiety or nervousness on the day of surgery. In a recall situation however, if patients can accurately recall their prior health state, then random error should be the only source of error. If these assumptions are true, then we might expect the agreement in a recall situation to be higher than that observed in a test re-test situation. Our results however, show that in fact the opposite is observed, suggesting that there is an additional source of error as a result of asking people to recall.

Finally, while the majority of studies that investigate the accuracy of a patient's ability to recall preoperative health status did so for the purpose of conducting unplanned, retrospective studies, our purpose was to determine whether we could plan in advance to collect recalled ratings of quality of life, general health, and functional status in a prospective randomized trial, to improve the efficiency of data collection. Because "recall error" is present, it is important to investigate the effect of this error on sample size requirements or the power to make statistical comparisons at the end of the study. It was expected that recalled ratings would increase within-subject error, leading to a greater overall error, or variance, leading to larger estimates of sample size or a reduction in statistical power (increase in Type II error rate). These expectations were confirmed by our results; recalled ratings did have greater associated variances (Table 7). Therefore, if planning to use recall ratings in place of prospectively collected baseline data, researchers must decide whether the gains in efficiency through data collection (i.e. reduction in patient burden and research staff resources at the front end of the study) are worth the increases in estimates of sample size or loss of power to make statistical comparisons at the study's conclusion. For example, in studies involving patients with rare diseases, it may be more feasible to expend resources in collecting baseline data prospectively than to require a greater number of patients in the study.

Finally, our findings highlight an important point about how analyses are conducted in clinical trials. Specifically, we commonly see three statistical approaches to between-group comparisons including, 1) a t test of the postoperative final outcome scores, 2) a t-test of the change score in which the preoperative end point is subtracted from the postoperative measurement, or 3) an analysis of covariance (ANCOVA), in which the baseline preoperative measure provides a covariate and the postoperative end point serves as the dependent variable. When conducting an analysis involving the use of either a change score or an ANCOVA, the magnitude of the association between the preoperative baseline and postoperative end point scores is extremely important. The use of postoperative end point scores has been described as inefficient by several authors if the magnitude of the correlation between baseline and final postoperative ratings is greater than 0[63-68]. For an analysis involving the use of change scores, the magnitude of this correlation needs to be at least 0.5 to avoid losses in statistical power[65-67]. For an ANCOVA, statistical power is greater than that achieved using a change score or post-only score as soon as the correlation between pre- and post scores is greater than 0 and increases as the strength of the association increases[63,65,67].

Only two of the questionnaires in our study (SF-12 PCS, Feeling thermometer) had a correlation greater than 0.5 between pre- and post-intervention scores suggesting that a loss of power is probable if investigators are using change scores to evaluate outcomes. Moreover, since the correlation between pre- and post- scores is greater than 0 across all questionnaires, comparisons using post-only scores will also have less power than comparisons using an ANCOVA[63-67,69,70].

The results of the present study demonstrate that use of ANCOVA provides considerable advantage over change scores and post-only scores for between-group comparisons (Table 7). A further benefit is that if researchers plan to use recall data in place of prospectively collected baseline data using an ANCOVA, but do not have sufficient data to estimate the increase in variance or the strength of the correlation between pretest and posttest data, a conservative estimate for sample size can be obtained by using calculations meant for post-only comparisons.

Strengths of this study are that it was a large randomized controlled trial with multiple surgeons participating in recruitment, with a wide variety of self-assessment instruments used to assess outcome (hip-specific, disease-specific, and generic health measures). Our sample included patients with varying severities of osteoarthritis ranging from mild to severe, and undergoing either a primary or a revision total hip arthroplasty. This is the first study to investigate recall specifically among senior patients, with an average age of study participants of 71 years (range 55 to 90), making the results generalizable to an elderly population.

## Chapter 7: Conclusion

Patients undergoing total hip arthroplasty can recall their preoperative quality of life, general health, and functional status at 6 weeks postoperatively with sufficient accuracy to warrant substituting prospectively collected baseline data with retrospective ratings. These results suggest that investigators can improve the efficiency of data collection for randomized clinical trials in this patient population when investigating changes following an intervention at the group level, with minimal expected loss of statistical power, given the use of an efficient statistical test.

We recommend that retrospective baseline data collection be used in clinical trials in which patient eligibility is not known for certain prior to surgical evaluation, and the potential for post-surgical exclusion is high.

## References

1. Health Canada. Arthritis in Canada. An Ongoing Challenge. 2003 2003;H39-4/14-2003E.

2. Canadian Institute for Health Information (CIHI). Canadian Joint Replacement registry, 2007 Annual Report: Hip and Knee Replacements in Canada. 2006;ISBN 978-1-55465-189-4.

3. Bryant D, Norman G, Stratford P, Marx RG, Walter SD, Guyatt G. Patients undergoing knee surgery provided accurate ratings of preoperative quality of life and function 2 weeks after surgery. *J Clin Epidemiol* 2006 Sep;59(9):984-93.

4. Kang MN, Berry DJ, Maloney III, William J. Lower Extremity Considerations: Hip. In: Moskowitz RW, editor. Osteoarthritis: diagnosis and medical/surgical management. 4th ed. ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2007. p.375-93.

5. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br* 2007 Aug;89(8):1010-4.

6. Zimmer Inc. Hip Replacement. 2006 March 3, 2006;2008(January 10):1.

7. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996 Mar;78(2):185-90.

8. Singer AJ, Kowalska A, Thode HC,Jr. Ability of patients to accurately recall the severity of acute painful events. *Acad Emerg Med* 2001 Mar;8(3):292-5.

9. Babul N, Darke AC, Johnson DH, Charron-Vincent K. Using memory for pain in analgesic research. *Ann Pharmacother* 1993 Jan;27(1):9-12.

10. Beese A, Morley S. Memory for acute pain experience is specifically inaccurate but generally reliable. *Pain* 1993 May;53(2):183-9.

11. Zonneveld LN, McGrath PJ, Reid GJ, Sorbi MJ. Accuracy of children's pain memories. *Pain* 1997 Jul;71(3):297-302.

12. Hunter M, Philips C, Rachman S. Memory for pain. *Pain* 1979 Feb;6(1):35-46.

13. Brauer C, Thomsen JF, Loft IP, Mikkelsen S. Can we rely on retrospective pain assessments? *Am J Epidemiol* 2003 Mar 15;157(6):552-7.

14. Kreulen GJ, Stommel M, Gutek BA, Burns LR, Braden CJ. Utility of retrospective pretest ratings of patient satisfaction with health status. *Res Nurs Health* 2002 Jun;25(3):233-41.

15. Pellise F, Vidal X, Hernandez A, Cedraschi C, Bago J, Villanueva C. Reliability of retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low back pain. *Spine* 2005 Feb 1;30(3):365-8.

16. Dawson EG, Kanim LE, Sra P, Dorey FJ, Goldstein TB, Delamarter RB, et al. Low back pain recollection versus concurrent accounts: outcomes analysis. *Spine* 2002 May 1;27(9):984,93; discussion 994.

17. Feine JS, Lavigne GJ, Dao TT, Morin C, Lund JP. Memories of chronic pain and perceptions of relief. *Pain* 1998 Aug;77(2):137-41.

18. Herrmann D. Reporting current, past, and changed health status. What we know about distortion. *Med Care* 1995 Apr;33(4 Suppl):AS89-94.

19. Bryant RA. Memory for pain and affect in chronic pain patients. *Pain* 1993 Sep;54(3):347-51.

20. Everts B, Karlson B, Wahrborg P, Abdon N, Herlitz J, Hedner T. Pain recollection after chest pain of cardiac origin. *Cardiology* 1999;92(2):115-20.

21. Jamison RN, Sbrocco T, Parris WC. The influence of physical and psychosocial factors on accuracy of memory for pain in chronic pain patients. *Pain* 1989 Jun;37(3):289-94.

22. ten Klooster PM, Drossaers-Bakker KW, Taal E, van de Laar MA. Can we assess baseline pain and global health retrospectively? *Clin Exp Rheumatol* 2007 Mar-Apr;25(2):176-81.

23. Aseltine RH,Jr, Carlson KJ, Fowler FJ,Jr, Barry MJ. Comparing prospective and retrospective measures of treatment outcomes. *Med Care* 1995 Apr;33(4 Suppl):AS67-76.

24. Lingard EA, Wright EA, Sledge CB, Kinemax Outcomes Group. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. *J Bone Joint Surg Am* 2001 Aug;83-A(8):1149-56.

25. Mancuso CA, Charlson ME. Does recollection error threaten the validity of cross-sectional studies of effectiveness? *Med Care* 1995 Apr;33(4 Suppl):AS77-88.

26. Linton SJ, Melin L. The accuracy of remembering chronic pain. *Pain* 1982 Jul;13(3):281-5.

27. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med* 1999 Jun;48(11):1531-48.

28. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med* 1999 Jun;48(11):1507-15.

29. Ross M. Relation of Implicit Theories to the Construction of Personal Histories. *Psychological Review* 1989;96(2):341-57.

30. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. *Qual Life Res* 2003 May;12(3):239-49.

31. Baddeley A. Human Memory. Theory and Practice. Revised Edition. Needham Heights, MA: Allyn and Bacon; 1998.

32. Bernhard J, Lowy A, Maibach R, Hurny C, Swiss Group for Clinical Cancer Research (SAKK) and the Swiss Institute for Applied Cancer Research (SIAK), Bern, Switzerland. Response shift in the perception of health for utility evaluation. an explorative investigation. *Eur J Cancer* 2001 Sep;37(14):1729-35.

33. Festinger L. A theory of social comparison processes. *Hum Relat* 1954;7:117.

34. Gedney JJ, Logan H, Baron RS. Predictors of short-term and long-term memory of sensory and affective dimensions of pain. *J Pain* 2003 Mar;4(2):47-55.

35. Kent G. Memory of dental experiences as related to naturally occurring changes in state anxiety. *Cognitive Emotion* 1989;3:45-53.

36. Kent G. Memory of dental pain. *Pain* 1985;21:187-94.

37. Eich E, Reeves JL, Jaeger B, Graff-Radford SB. Memory for pain: relation between past and present pain intensity. *Pain* 1985 Dec;23(4):375-80.

38. Elliott AM, Smith BH, Hannaford PC, Smith WC, Chambers WA. Assessing change in chronic pain severity: the chronic pain grade compared with retrospective perceptions. *Br J Gen Pract* 2002 Apr;52(477):269-74.

39. Moller JT, Cluitmans P, Rasmussen LS, Houx P, Rasmussen H, Canet J, et al. Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. International Study of Post-Operative Cognitive Dysfunction. *Lancet* 1998 Mar 21;351(9106):857-61.

40. BEDFORD PD. Adverse cerebral effects of anaesthesia on old people. *Lancet* 1955 Aug 6;269(6884):259-63.

41. Jones MJ, Piggott SE, Vaughan RS, Bayer AJ, Newcombe RG, Twining TC, et al. Cognitive and functional competence after anaesthesia in patients aged over 60: controlled trial of general and regional anaesthesia for elective hip or knee replacement. *BMJ* 1990 Jun 30;300(6741):1683-7.

42. Rasmussen LS, Johnson T, Kuipers HM, Kristensen D, Siersma VD, Vila P, et al. Does anaesthesia cause postoperative cognitive dysfunction? A randomised study of

regional versus general anaesthesia in 438 elderly patients. *Acta Anaesthesiol Scand* 2003 Mar;47(3):260-6.

43. Williams-Russo P, Sharrock NE, Mattis S, Szatrowski TP, Charlson ME. Cognitive effects after epidural vs general anesthesia in older adults. A randomized trial. *JAMA* 1995 Jul 5;274(1):44-50.

44. Maurer SG, Chen AL, Hiebert R, Pereira GC, Di Cesare PE. Comparison of outcomes of using spinal versus general anesthesia in total hip arthroplasty. *Am J Orthop* 2007 Jul;36(7):E101-6.

45. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Phys Ther* 1999 Apr;79(4):371-83.

46. Watson CJ, Propps M, Ratner J, Zeigler DL, Horton P, Smith SS. Reliability and responsiveness of the lower extremity functional scale and the anterior knee pain scale in patients with anterior knee pain. *J Orthop Sports Phys Ther* 2005 Mar;35(3):136-46.

47. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988 Dec;15(12):1833-40.

48. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the western Ontario and McMaster Universities Osteoarthritis Index and the

short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. *Arthritis Care Res* 1999 Jun;12(3):172-9.

49. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 2000 Nov;27(11):2635-41.

50. Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg Br* 2001 Nov;83(8):1125-9.

51. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Health Care* 1996 Jun;5(2):81-8.

52. Pynsent PB, Adams DJ, Disney SP. The Oxford hip and knee outcome questionnaires for arthroplasty. *J Bone Joint Surg Br* 2005 Feb;87(2):241-8.

53. Ware J,Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996 Mar;34(3):220-33.

54. Puhan MA, Guyatt GH, Montori VM, Bhandari M, Devereaux PJ, Griffith L, et al. The standard gamble demonstrated lower reliability than the feeling thermometer. *J Clin Epidemiol* 2005 May;58(5):458-65.

55. Schunemann HJ, Griffith L, Stubbing D, Goldstein R, Guyatt GH. A clinical trial to evaluate the measurement properties of 2 direct preference instruments administered with and without hypothetical marker states. *Med Decis Making* 2003 Mar-Apr;23(2):140-9.

56. Schunemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbing D, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *J Clin Epidemiol* 2003 Dec;56(12):1170-6.

57. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002 May 15;21(9):1331-5.

58. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther* 1997 Jul;77(7):745-50.

59. Pitman EJG. A note on normal correlation. *Biometrika* 1939;31:9-12.

60. Snedecor GW, Cochran WG. Statistical methods. 6th ed. ed. Ames, Iowa: Iowa State University Press; 1967.

61. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. *J Rheumatol* 1993 Mar;20(3):561-5.

62. Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthritis Cartilage* 2005 Dec;13(12):1076-83.

63. Cronbach LJ, Furby L. How should we measure 'change' - or should we? *Psychol Bull* 1970;74:68-80.

64. Knapp TR. The (un)reliability of change scores in counseling research. *Meas Eval Guid* 1980;13:149-57.

65. Lee J. A note on the comparison of group means based on repeated measurements of the same subject. *J Chronic Dis* 1980;33(10):673-5.

66. Norman GR. Issues in the use of change scores in randomized trials. *J Clin Epidemiol* 1989;42:1097-105.

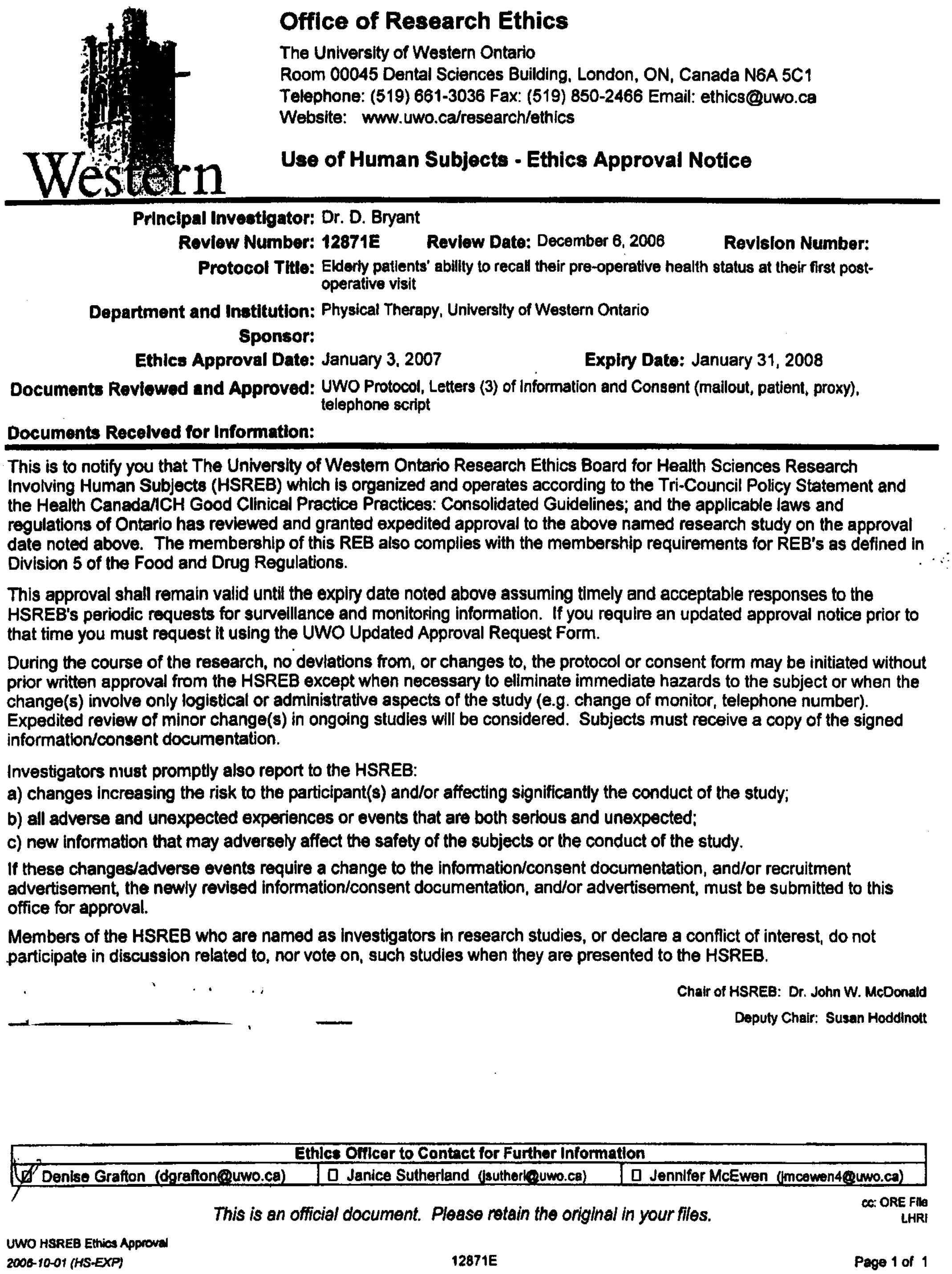
67. Oldham PD. A note on the analysis of repeated measurements of the same subjects. *J Chronic Dis* 1962;15:969-77.

68. Stanek EJ. Choosing a pretest-posttest analysis. *Am Stat* 1988;42:178-83.

69. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med* 1992 Sep 30;11(13):1685-704.

70. Egger MJ, Coleman ML, Ward JR, Reading JC, Williams HJ. Uses and abuses of analysis of covariance in clinical trials. *Control Clin Trials* 1985 Mar;6(1):12-24.

# Appendix A: Ethics Approval



Western

**Office of Research Ethics**

The University of Western Ontario
Room 00045 Dental Sciences Building, London, ON, Canada N6A 5C1
Telephone: (519) 661-3036 Fax: (519) 850-2466 Email: ethics@uwo.ca
Website: www.uwo.ca/research/ethics

**Use of Human Subjects - Ethics Approval Notice**

**Principal Investigator:** Dr. D. Bryant
**Review Number:** 12871E **Review Date:** December 6, 2006 **Revision Number:**
**Protocol Title:** Elderly patients' ability to recall their pre-operative health status at their first post-operative visit
**Department and Institution:** Physical Therapy, University of Western Ontario
**Sponsor:**
**Ethics Approval Date:** January 3, 2007 **Expiry Date:** January 31, 2008
**Documents Reviewed and Approved:** UWO Protocol, Letters (3) of Information and Consent (mailout, patient, proxy), telephone script
**Documents Received for Information:**

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted expedited approval to the above named research study on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

This approval shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

Chair of HSREB: Dr. John W. McDonald
Deputy Chair: Susan Hoddinott

| Ethics Officer to Contact for Further Information | | |
|---|---|---|
| ☑ Denise Grafton (dgrafton@uwo.ca) | □ Janice Sutherland (jsutherl@uwo.ca) | □ Jennifer McEwen (jmcewen4@uwo.ca) |

*This is an official document. Please retain the original in your files.*

cc: ORE File
LHRI

UWO HSREB Ethics Approval
*2006-10-01 (HS-EXP)*

12871E

Page 1 of 1