Response Shift Following Surgery of the Lumbar Spine

By

Joel A. Finkelstein MD, FRCS(C)

A thesis submitted in
conformity with the
requirements

for the degree of Master of Science

Graduate Department:
Institute of Medical Sciences

University of Toronto

© Copyright by Joel Alan Finkelstein 2010
Abstract

This study is a prospective longitudinal outcome study investigating the presence of response shift in disease and generic functional outcome measures in 105 patients undergoing spinal surgery. The then-test method which compares pre-test scores to retrospective pre-test scores was used to quantitate response shift. There was a statistically significant response shift for the Oswestry Disability Index (ODI) ($p=0.001$) and the Short Form-36-PCS ($p=0.078$). At three months, seventy-two percent of patients exhibited a response shift with the ODI. Fifty-six and 21 percent of patients exhibited a response shift with the SF-36 physical and mental component scores respectively. When accounting for response shift and using the minimal clinically important difference, the success rate of the surgery at 3 months increased by 20 percent. The presence of response shift has implications for the measurement properties of standard spinal surgery outcome measures including the effect size of treatment and the number of responders to treatment.
Acknowledgements

I would like to acknowledge the Institute of Medical Science for the opportunity to complete this thesis. Special extended acknowledgement is given to Aileen Davis, who put up with endless edits and provided unparalleled guidance to complete this project. I would like to thank my committee members, Gary Naglie and Alex Kiss and to Carolyn Gimera for her assistance with the preparation and formatting of the manuscript. Finally, I appreciate the professional support given by the Department of Surgery University of Toronto and the Division of Orthopaedic Surgery of Sunnybrook Health Sciences Center.
<table>
<thead>
<tr>
<th>Table of Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>xiii</td>
</tr>
<tr>
<td>Preface</td>
<td>xiv</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION AND BACKGROUND

1.0. INTRODUCTION

1.1. Response Shift Illustration
1.2. Prospective vs. Retrospective Measures of Change
1.3. Theoretical Model of Response Shift

2.0. RESPONSE SHIFT DETECTION METHODS

2.1. Individual Patient Data Collection Methods
2.1.1 Then - Test
2.1.2 Other Individual Patient Data Methods
2.2. Statistical Approaches
2.2.1 Structural Equation Modeling
2.2.2 Latent Trajectory Analysis with Subject-Centered Residuals
2.2.3 Classification and Regression Tree Analysis (CART)
2.2.4 Differential Item Function (DIF) Analysis
2.3. Comparing Methods of Detecting Response Shift

3.0. RESPONSE SHIFT AND BIAS

3.1. Recall Bias
3.2 Motivational Bias and Effort Justification Bias

4.0. RESPONSE SHIFT VS. IMPLICIT THEORY OF CHANGE

5.0. QUALITY OF LIFE AS A DYNAMIC CONSTRUCT

5.1. Adaptation
5.2. Response Shift as an Adaptive Mechanism 13
5.3. Adaptation Studies on Response Shift 14
5.4. Antecedents and Mechanisms for Response Shift 15

6.0. DEGENERATIVE SPINAL CONDITIONS 16
6.1. Scope of the Problem and Treatment Overview 16
6.2. Spinal Surgery and Response Shift 17
6.3. Functional Outcome Measures in Surgery 17
   6.3.1 The International Classification of Functioning, Disability and Health 18
6.4. Disease-Specific Functional Outcome Measures 18
   6.4.1. Oswestry Disability Index 18
   6.4.2. Oswestry Disability Index Measurement Properties 19
      6.4.2.1 Face and Content Validity 19
      6.4.2.2 Construct Validity 19
      6.4.2.3 Internal Consistency 20
   6.4.4.4 Reliability 20
   6.4.4.5 Responsiveness 20
   6.4.4.6 Minimal Clinically Important Difference 20
6.5. Generic Outcome Measures in Spinal Disorders 21
   6.5.1 Short Form 36 21
      6.5.1.1 Reliability 22
      6.5.1.2 Validity 22
      6.5.1.3 Minimal Clinically Important Difference 22
CHAPTER 2: OUTCOME FOLLOWING SPINAL SURGERY: THE INFLUENCE OF RESPONSE SHIFT

1.0. ABSTRACT

2.0. INTRODUCTION

3.0. METHODS
   3.1 Study Design and Setting
   3.2 Participants
   3.3 Data Collected
      3.3.1 Primary and Secondary Outcomes
      3.3.2 Descriptive Data
      3.3.3 Potential Predictor Variables of Response Shift

4.0 PROCEDURES

5.0 ANALYSIS
   5.1 Minimum Detectable Change
   5.2 Minimal Clinically Important Difference
<table>
<thead>
<tr>
<th>Section</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3 Consistency of Response Shift</td>
<td>35</td>
</tr>
<tr>
<td>5.4 Predictor Variables</td>
<td>35</td>
</tr>
<tr>
<td>5.5 Effect Size</td>
<td>36</td>
</tr>
<tr>
<td>5.6 Sample Size Calculation</td>
<td>36</td>
</tr>
<tr>
<td>6.0 RESULTS</td>
<td>36</td>
</tr>
<tr>
<td>6.1 Patient Accrual</td>
<td>36</td>
</tr>
<tr>
<td>6.2 Out of Study Eligible Patients Data</td>
<td>37</td>
</tr>
<tr>
<td>6.3 Outcome Scores</td>
<td>39</td>
</tr>
<tr>
<td>6.4 Post-hoc Power Calculation</td>
<td>41</td>
</tr>
<tr>
<td>6.5 Frequency and Direction of Response Shift</td>
<td>42</td>
</tr>
<tr>
<td>6.6 Consistency of Response Shift at 6 Weeks and 3 Months</td>
<td>44</td>
</tr>
<tr>
<td>6.7 Minimal Clinically Important Difference</td>
<td>46</td>
</tr>
<tr>
<td>6.8 Predictor Variables</td>
<td>47</td>
</tr>
<tr>
<td>7.0 DISCUSSION</td>
<td>48</td>
</tr>
<tr>
<td>7.1 Primary and Secondary Objectives</td>
<td>48</td>
</tr>
<tr>
<td>7.2 Limitations and Strengths</td>
<td>53</td>
</tr>
<tr>
<td>7.3 Future Directions</td>
<td>54</td>
</tr>
<tr>
<td>CHAPTER THREE: THESIS DISCUSSION</td>
<td>56</td>
</tr>
<tr>
<td>1.0 Thesis Discussion</td>
<td>56</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>59</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>72</td>
</tr>
</tbody>
</table>
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODI</td>
<td>Oswestry Disability Index</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short Form-36</td>
</tr>
<tr>
<td>SF-36-MCS</td>
<td>Short Form-36- Mental Component Score</td>
</tr>
<tr>
<td>SF-36-PCS</td>
<td>Short Form-36 - Physical Component Score</td>
</tr>
<tr>
<td>SF-12</td>
<td>Short Form-12</td>
</tr>
<tr>
<td>QoL</td>
<td>Quality of Life</td>
</tr>
<tr>
<td>MYMOP</td>
<td>Measure Yourself Medical Outcome Profile</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>SEIQOL</td>
<td>Schedule for the Evaluation of Individual Quality of Life</td>
</tr>
<tr>
<td>SEM</td>
<td>Structural Equation Modelling</td>
</tr>
<tr>
<td>CART</td>
<td>Classification and Regression Tree Analysis</td>
</tr>
<tr>
<td>DIF</td>
<td>Differential Item Functioning</td>
</tr>
<tr>
<td>RS</td>
<td>Response Shift</td>
</tr>
<tr>
<td>EORTC QLC</td>
<td>The European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire</td>
</tr>
<tr>
<td>SPORT</td>
<td>Spine OutcomeResearch Trial</td>
</tr>
<tr>
<td>ICF</td>
<td>The International Classification of Functioning, Disability, and Health</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>MCID</td>
<td>Minimal Clinically Important Difference</td>
</tr>
<tr>
<td>MDC</td>
<td>Minimal Detectable Change</td>
</tr>
<tr>
<td>LISAT-11</td>
<td>Life Satisfaction Questionnaire</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analogue Scale</td>
</tr>
<tr>
<td>REB</td>
<td>Research Ethics Board</td>
</tr>
<tr>
<td>TE</td>
<td>Treatment Effect</td>
</tr>
<tr>
<td>ATE</td>
<td>Adjusted Treatment Effect</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard Error of the Measurement Instrument</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Sqrt</td>
<td>Square root</td>
</tr>
<tr>
<td>r</td>
<td>Test–retest Reliability Coefficient</td>
</tr>
<tr>
<td>ES</td>
<td>Effect Size</td>
</tr>
<tr>
<td>WOMAC</td>
<td>Western Ontario and McMaster Universities Osteoarthritis Index</td>
</tr>
<tr>
<td>List of Figures</td>
<td>Page No.</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Figure 1. Response Shift Model</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2. Then – Test Design</td>
<td>5</td>
</tr>
<tr>
<td>Figure 3. Patient Accrual</td>
<td>37</td>
</tr>
<tr>
<td>Table No.</td>
<td>Title</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Table 1</td>
<td>Summary Equations</td>
</tr>
<tr>
<td>Table 2</td>
<td>Minimal Detectable Changes, Standard Deviations and Test-retest Reliability Coefficients of Outcome Measures</td>
</tr>
<tr>
<td>Table 3</td>
<td>Comparative Data for the In-Study vs. Out of Study Patients</td>
</tr>
<tr>
<td>Table 4</td>
<td>Characteristics of Patients with One vs. Two Follow-up Assessments</td>
</tr>
<tr>
<td>Table 5</td>
<td>Most Common Co-morbid Conditions</td>
</tr>
<tr>
<td>Table 6</td>
<td>6 Week Summary Scores</td>
</tr>
<tr>
<td>Table 7</td>
<td>3 Month Summary Scores</td>
</tr>
<tr>
<td>Table 8</td>
<td>Adjusted and Unadjusted Treatment Effects and Effect Size</td>
</tr>
<tr>
<td>Table 9</td>
<td>Direction of Response Shift: ODI</td>
</tr>
<tr>
<td>Table 10</td>
<td>Direction of Response Shift: SF-36-PCS</td>
</tr>
<tr>
<td>Table 11</td>
<td>Direction of Response Shift: SF-36-MCS</td>
</tr>
<tr>
<td>Table 12</td>
<td>ODI Consistency of Response Shift Direction at 6 Weeks and 3 Months</td>
</tr>
<tr>
<td>Table 13</td>
<td>SF-36-PCS Consistency of Response Shift Direction at 6 Weeks and 3 Months</td>
</tr>
<tr>
<td>Table 14</td>
<td>Minimal Clinically Important Difference – 3 Month ODI</td>
</tr>
<tr>
<td>Table 15</td>
<td>Comparison of Means for Potential Predictor Variables</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>Page No.</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>1. Oswestry Disability Index</td>
<td>72</td>
</tr>
<tr>
<td>2. SF- 36</td>
<td>74</td>
</tr>
<tr>
<td>3. Research Ethics Board Approval</td>
<td>79</td>
</tr>
<tr>
<td>4. LISAT -11</td>
<td>82</td>
</tr>
</tbody>
</table>
Preface

Patient self-ratings of health-related quality of life (QoL), general health and functional status are important outcome measures when evaluating the effect of clinical interventions.[1] Validated self-report outcome questionnaires are the basis for measuring changes in a patient's perspective of QoL and in their function after a given intervention. A challenge to the interpretation of prospective evaluation of change, pre-post scores, may be that when patients experience changes in health, they can “adapt” and change their internal standards, their values, and their conceptualization of key aspects of their illness. As such, these prospective self-report measures are subject to changing reference values or adaptation. These “response shifts” may have effects that conceal treatment effects over time.

Examples of patient adaptation exist with paradoxical findings of cancer patients reporting comparable QoL to healthy controls, or with patients rating their QoL to be better than their QoL as rated by their health care providers.[2, 3] This has been called “recalibration response shift”. This results in a change in the reference value by which patients evaluate their QoL over time.[4] Ideally, measurement of outcome needs to be patient centered, able to capture the patient’s perspective of change and be able to accommodate to changes in QoL appraisal or recalibration over time.[5-7]

This thesis begins by providing a rationale of why response shift is important in outcome measurement. It then describes the theoretical models of response shift, discusses methods of evaluating response shift and possible mechanisms for the phenomenon. This information provides the background for the research presented in the manuscript format that formed the basis of this thesis. The thesis closes with a discussion of the research results in the context of the current literature and knowledge.
CHAPTER 1

INTRODUCTION AND BACKGROUND

1.0 Introduction

Response shift theory suggests that changes in individual's health status may produce behavioral, cognitive and affective changes that can alter an individual's internal standards, values or conceptualization of health related quality of life (QoL).[4, 8] Response shift occurs after treatment and is a re-evaluation of one’s pre-treatment state that includes QoL, pain and disability. Response shift can occur in any situation where self-reports are used to evaluate treatment outcomes.[9] Stated another way, response shift implies that people may shift their goalposts over time so that even if a condition remains unchanged, it is scored differently by the same self-report questionnaire.

1.1 Response Shift Illustration

An illustration of response shift and its effect on outcome measurement is provided by Paterson.[10] In measuring Euroqol scores and the Measure Yourself Medical Outcome Profile (MYMOP) over a period of 6 months in patients undergoing acupuncture for chronic disorders, a patient describes how, on the first occasion, she scored her daily/usual activities in relation to her inability to do her paid work. But after she had to give up her paid work, she scored them only in relation to her ability to do her household tasks. Consequently the scores improved, although her functional ability was unchanged.

Response shift can form a valuable strategy for coping with the reality of a chronic disease by recalibrating one’s expectations for health and functioning and also recalibrating the relative valuation of health states.[11] In another illustration of response shift, Schwartz et al. described in a cohort of young adult cancer survivors, how after a psychosocial intervention there seemed to be a detrimental impact on measured aspects of well-being when using standard pre-post
A closer look at the results, however, indicated that the young adult survivors of cancer shifted in how physical functioning was valued and this reappraisal of patients’ values contributed to their global reassessment of QoL. After the intervention, these patients placed more value on the physical functions that they could not do before and they valued these physical parameters in a manner more similar to healthy individuals. Thus, the intervention was indeed successful because the cancer survivors were now functioning more similarly to healthy peers, but when reexamining their QoL pre-treatment, they now rated it worse than they had before the intervention.

1.2 Prospective vs. Retrospective Measures of Change

The preceding examples illustrate how response shift can obfuscate treatment effects over time. Adaptation to changed circumstances can result in a change in the reference value by which patients evaluate their QoL. Such change can create a bias in longitudinal comparisons of QoL since these assume a constant reference value.[4, 9, 13]

In order to accommodate for a patient’s change in reference, measurement of response shift has used a technique that assesses current reference values in a retrospective format in order to assess baseline status; this is the then-test. The then-test, (also called a retrospective pre-test) is administered at follow-up, asking patients to provide a renewed judgment about their pre-treatment QoL. Patients also are asked to concurrently assess their outcome post-treatment. Both measures are performed at the same time, thus assuming that the same internal reference value is used for both measurements. The retrospective pre-test compared to the follow-up measurement has been proposed as a method to assess change in QoL or outcome of a given treatment intervention that is not confounded by change in reference values.[14]

The usefulness of the then-test was noted by Nieuwkerk et al. in a study where they compared change in QoL in Human Immunodeficiency Virus (HIV) positive patients who were started on a retroviral therapy regimen.[15] Pre-post change after treatment was measured using conventional baseline and follow-up measurements as well as retrospective pre-test (then-test) and follow-up measurements. The then-test approach to measurement showed larger change in clinical
indicators of health status, suggesting a more valid measurement of change than with a conventional prospective method.

1.3 Theoretical Model of Response Shift

A theoretical model of response shift has been proposed by Sprangers and Schwartz, 1999. (Figure 1)[13]

Figure 1. Response Shift Model

This model proposes that stable personality characteristics (i.e., antecedents) interact with cognitive, affective, or behavioral processes that individuals use to deal with life changes (i.e., mechanisms). These promote response shifts that then result in a level of perceived QoL that is higher or lower than expected based upon objective criteria. The response shift process is hypothesized to be happening continuously as part of a positive or negative adaptation process.
Rapkin and Schwartz expanded upon this model by adding an appraisal profile to the model. The Appraisal Profile involves cognitive debriefing of the patient that is useful for understanding the cognitive processes used by the patient that lead to a response shift.[7]

2.0 Response-Shift Detection Methods

There are a number of methods for detecting response shift. These include individualized methods, preference-based methods, qualitative methods, and statistical methods.[4] These methods can be broadly conceptualized as either: (1) those that require response-shift-specific data collection; and, (2) those that rely on secondary analysis of existing data to detect response shift.

2.1 Individualized Patient Data Collection Methods

2.1.1 Then-Test

The then-test approach has been introduced above. Howard et al were the first to implement a then-test methodology; also described as the retrospective pre-test design approach.[16] In this approach, respondents complete pre-test questionnaires at baseline (pre-intervention) and post-test questionnaires at follow-up, and additionally they are asked at the time of completion of the post-test to retrospectively evaluate their health status or QoL as they perceived it to be at the pre-treatment time (i.e. the then-test). If the then-test is completed at the same time as the concurrent follow-up measurement, it is assumed that the same reference value is used for both measurements.[14] Comparing the then-test with a follow-up measurement will assess change in QoL over time that is not confounded by change in reference values.[15]

Using the then-test, a patient's retrospective re-evaluation of their baseline status can be worse or better compared to their prospective baseline assessment.[11, 17-20]

The amount of response shift is the difference between the pre-test and the retrospective pre-test scores. The conventional post-test score minus the pre-test score is the routinely measured
change following treatment. This is the conventional treatment effect. The post-test score minus the then-test score is the adjusted treatment effect, which accounts for the response shift.

**Figure 2: Then Test Design**

<table>
<thead>
<tr>
<th>Then Test</th>
<th>Pre Test</th>
<th>Post Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response Shift Magnitude</th>
<th>Traditional pre/post test evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Treatment Effect)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adjusted Treatment Effect</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$T_{1r}$</th>
<th>$T_o$</th>
<th>$T_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>retrospective pre test</td>
<td>pre test (pre-operative)</td>
<td>post test</td>
</tr>
</tbody>
</table>

This design approach for detecting response shift has been widely used; however it is felt that recall bias may be a confounding factor.[21-24] This is addressed in Section 3.0.

Validity of the then-test has been established in educational research where the majority of the initial studies utilizing this approach were performed.[16] The then-test was found to be strongly associated with objective criteria of change which suggested it was an accurate and sensitive assessment of the respondent’s perspective of personal change.[8] The reliability of the then-test approach is equivalent to the reliability of the outcome measure being used.[14] A recent meta-analysis of response shift studies using the then-test to measure the effect size of moderator variables on an individual’s QoL indicated that the detected effect sizes were small to moderate depending on the measurement sample. Response shift can be meaningful and clinically important when used as a means of adjusting outcome study results.[25]
2.1.2 Other Individual Patient Data Methods

There are two further methods that have developed substantially in the past several years that use patient specific data: The Appraisal Profile of Rapkin[7] which combines qualitative and statistical data and the Schedule for the Evaluation of Individual Quality of Life (SEIQOL) [27] which is preference-based.

1) The Appraisal Profile of Rapkin and Schwartz uses a semi-structured interview format to assess what parameters a patient uses to assess his/her QoL.[7] This yields qualitative textual data in response to open-ended questions as well as quantitative data in response to multiple choice questions. The tool measures four distinct parameters in the appraisal process: (a) *Framing*: i.e., what does QoL mean to the individual?; (b) *Sampling*: i.e., what relevant experiences does the person have?; (c) *Evaluating*: i.e., how do experiences compare to relevant standards?; and, (d) *Combining*: i.e., what is the relative importance of different experiences?

2) The Schedule for the Evaluation of Individual Quality of Life (SEIQOL) has been used to examine stability of domains and internal standards (termed cues and anchors) over time and across groups.[26] This method involves a semi-structured interview that elicits from the person the domains or areas s/he thinks about when considering QoL (“cues”), the response options that make the most sense to the person (“anchors”), and his/her level of functioning within each of these domains (“levels”). A regression model is used to generate a person’s predicted score. This individualized QoL tool has been found to be useful in understanding changes in individuals over time as they cope with a progressive and deteriorating disease. The SEIQOL approach has been successfully applied in a variety of patient populations, including cancer, hip replacement surgery and dental treatments.[27, 28]
2.2 **Statistical Approaches**

There are four statistical methods that have been applied to response shift detection. 1) Structural equation modeling; 2) latent trajectory analysis with subject-centered residuals; 3) classification and regression tree analysis; and; 4) differential item functioning. All of these methods require substantial sample sizes, i.e. minimum of 200 patients. These methods vary in terms of how much they focus on aggregate analyses versus individual patient-focused analyses, and thus how sensitive they are to individual-level response shift.

2.2.1 **Structural Equation Modeling**

Originally evolving from factor analytic methods, structural equation modeling (SEM) is a technique that combines factor analysis and regression analysis to solve multivariable research questions at a group level.[29] By analyzing covariance matrices, these models test measurement and structural models to first test the assumption of measurement invariance and then examine whether relationships among variables are similar over time (i.e., the structural model). Recent advances of this method were made by Oort and colleagues to clarify how distinct changes detectable with SEM reflect different aspects of response shift, i.e. change in internal standards, reconceptualization or recalibration.[30] SEM has the advantage of allowing secondary analyses of existing data to test response shift hypotheses. It has the disadvantage of being sensitive to response shifts only when a majority of the sample demonstrates a response shift.[29]

2.2.2 **Latent Trajectory Analysis with Subject-Centered Residuals**

This method seeks to develop a predictive growth curve model to examine patterns in discrepancies between expected and observed scores.[30] By obtaining and scaling model residuals, subject-centered residuals are created to categorize respondents as either: (1) exhibiting no response shift, i.e., the person’s residuals are consistent over time but there was some change in their perceived QoL; (2) exhibiting a positive response shift, i.e., the person’s evaluation started lower than expected and then shifted or reassessed upward; or,
(3) exhibiting a negative response shift, i.e., started higher than expected and then reassessed downward over time. This method is of interest because it classifies response shift at the individual rather than group level and because it distinguishes people based on the timing as well as the direction of the response shift. It is useful for stratified analyses with existing data and thus does not impose additional demands on the respondent. Its primary weakness is that it cannot distinguish random error from response shift. Like other statistically sophisticated methods, it requires a substantial sample size measured over multiple time points to create a predictive model.

2.2.3 Classification and Regression Tree Analysis (CART)

Classification tree (also known as decision tree) methods are useful for data mining in order to make prediction of outcomes and to generate rules that can be easily understood and explained. This method was applied by Li and Rapkin and combines qualitative and quantitative methods to yield patient data.[31] These investigators utilized the Appraisal Profile developed by Rapkin to yield qualitative textual data in response to open-ended questions. Multiple choice questions were used to obtain quantitative data.[7] These data were then content analyzed to yield categories amenable to quantitative analysis and “trees” were generated.[31] The classification tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions and then splitting it further on each of the branches. An algorithm then systematically breaks up the records into two parts, examining one variable at a time and splitting the records on the basis of a dividing line. This splitting or partitioning is then applied to a new category or domain. This process continues until no more useful splits can be found. The object is to attain a homogeneous set of patients in each partition and the final product of this analysis is groupings of respondents who share patterns of appraisal.

2.2.4 Differential Item Functioning (DIF) Analysis

DIF analysis is a statistical method that attempts to assess whether an item functions differently between groupings of people.[32] In the context of response shift research, DIF over time within subjects would examine whether time-related changes (in health or due to some catalytic experience such as surgery) are detectable in terms of changes in item functioning over time. In
this method, one controls for the overall (disability) score and analyzes (commonly using logistic
regression) how the same item in a questionnaire is used or may have a different meaning to
subgroups of people over the trajectory of their recovery. DIF can be used to identify response
shift if an individual changes how s/he measures an item in a questionnaire. Reconceptualization
DIF examines whether the groups conceptualize these items the same at each time point.
Recalibration DIF would examine differences over time between groups.

2.3 Comparing Methods of Detecting Response Shift

All of the methods described above have strengths and weaknesses. Whereas some are more
useful for group analyses, these methods may be limited if a substantial portion of the study
sample does not undergo response shifts. Individual-level methods may be more useful
clinically, especially if the clinician seeks to track a patient’s response shift and wishes to enable
such shifts as part of the healing trajectory. In the case of a deteriorating health status due to a
life threatening disease, enabling response shift is a desired outcome of treatment by facilitating
coping and accommodating to the disease.[11]

The then-test is the simplest approach and enables calculation of effect sizes of the outcome
variable while the Appraisal Profile uses individual-level data and is amenable to modeling that
yields insights into the cognitive processes that are used by the patient to rate his/her QoL.

3.0 Response Shift and Bias

Bias refers to ratings that depart systematically from true values.[33] Biases in subjective
measurement can arise from the respondent’s personality, from the way they perceive
questionnaires, or from particular circumstances of their illnesses. Bias also exists in the desire to
give socially desirable responses and from the way patients interpret questionnaire response
scales.[34-36]
3.1 Recall Bias

The concern about recall bias, e.g. patients is unable to recall their preoperative level of function or QoL has been noted above. This is a threat to the validity of the then-test.[21, 22, 24, 37] A number of studies however have suggested that recall bias may not to be a concern.[14, 23, 37-39]

Visser looked at convergent validity for three methods for measuring response shift: structural equation modeling (SEM), anchor recalibration, and the then-test design.[23] These authors concluded that there was convergence in the detection of response shift between the then-test and SEM approach but that there was divergence with the anchor calibration approach. Eight of nine subscales evaluated over a three month period demonstrated the presence (3 subscales) or absence (6 subscales) of response shift using the SEM and then-test approaches. However, for one subscale the response shift was in the opposite direction and the magnitude of the shift of the remaining two subscales was larger using SEM. Although there were similar patterns with the anchor-based method, two further subscales demonstrated response shift with this method. Given that the SEM approach does not require retrospective data and there is convergence of results between the then-test and SEM, it was felt that recall bias did not invalidate the then-test.

In a study of people undergoing total hip arthroplasty, Howell evaluated pre-operative recall at 3 days, 6 weeks and 3 months post surgery. There were statistically significant differences in five of seven measures at 3 days post-surgery, in one measure 6 weeks post surgery and in no measures at 3 months post surgery.[38] These data suggest that recall bias is a concern in the early post-operative period but that people can recall their pre-operative status at six weeks and three months post surgery.

In summary, the literature is inconclusive about whether recall bias is a concern, particularly at six weeks and 3 months following an intervention. As such, it cannot be conclusively stated that a retrospective measure such as the then-test is completely free of recall bias.
3.2 Motivational Bias and Effort Justification Bias

Another form of bias is motivational bias or effort justification bias. These imply that the burden of undergoing surgery and the greater attention by the treating physician (placebo effect of surgery) leads to an over-estimation of improvement. These biases may affect the retrospective pre–test score, (then –test) and overstate the pre operative disability, or may affect the post–score such to overemphasize the improvement. This is not response shift but illustrates the confounding bias that can affect outcome measurement.[40, 41] These biases are not unique to a study of response shift, but may be a factor in any pre-post outcome measure. A study comparing surgical and non-surgical management for the same condition would need to be performed to control for effort justification bias.

4.0 Response Shift vs. Implicit Theory of Change

A weakness in the then-test approach is that the results provide little information about the underlying cognitive appraisal processes that has lead to response shift. Response shift theory attempts to explain how people make assessments of health over time. Response shift has been invoked as an explanation for several paradoxes in the health literature, such as: 1) patients with chronic diseases rate QoL similarly to healthy individuals; 2) patients tend to rate their QoL higher than their caregivers; and, 3) discrepancies arise between objective measures of health and self-rated health.[42] In a further example of findings consistent with response shift, Linton and Melin in a study of a treatment program for chronic pain noted that patients recalled having experienced more pain than they had initially indicated. As a consequence of the new information they acquired in the program, they came to perceive that their baseline pain was much worse than they thought.[43]

Response shift theorizes that the old state is adjusted in the retrospective judgment on the basis of new information acquired in the interim. Furthermore, the present state differs from a previous state as a result of a new internal standard and that a retrospective judgment of the previous state as measured by the then-test, also changes because it is reassessed in light of the new standard.[4]
Another theory that may account for the accuracy of patient recall or how patients make assessments of health over time is the implicit theory of change.[42] Implicit theorists suggest that patients do not remember their initial state and instead extrapolate backwards from their present state. Implicit theory presumes that memory or recall of the pre-treatment state is poor so that the retrospective judgment of the initial state is reconstructed and the prospective judgment is more valid. According to the implicit theory, 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.[37]

5.0 Quality of Life as a Dynamic Construct

With subjective assessments, examples of within-subject QoL construct variation were noted by a number of authors (i.e. an individual changing the standards by which he/she assess his/her QoL).[44] Cassileth found the psychological well-being of a group of melanoma patients to be significantly better than a group of patients with other less severe dermatological disorders, and slightly better than a group of the general population.[45] Decker and Schultz found the life satisfaction of a group of spinal cord injury patients to be only slightly worse than that of controls.[46] Irwin reported the life satisfaction among cancer survivors to be better than that of controls.[47] Evans reported, in a series of studies in transplant recipients and hemodialysis patients, that their level of happiness, satisfaction or QoL often exceeded that of a healthy population.[48]

The examples above illustrate what an external eye may view as a poorer QoL given that one’s personal reported QoL is based on his/her own internal standards, conceptualizations and values. These may change based on life experiences and may require adaptation to a new health status.

5.1 Adaptation

Adaptation is a psychological process by which past and present situations are given cognitive and emotional meaning so that an acceptable level of well-being is achieved.[49] De Haes and
Van Kippenbeg suggested that adaptation and downward social comparison might account for these paradoxical observations.[50] Chamberlain and Zika suggested that patients suffering from chronic diseases could experience an improvement in mood upon learning about others who are worse off. These individuals can adapt better to their disability by using a different scale of measurement that is more relevant to their present life status.[51] These changes in priorities and changes in groups of social comparison may create response shift that facilitates adaptation to the change in health status.

5.2 Response Shift as an Adaptive Mechanism

Another description of response shift is provided by Sprangers and Schwartz.[13] In their account, “an orthopaedic surgeon tells of a woman who after hearing her diagnosis of osteosarcoma, states that she would want euthanasia should her illness prevent her from walking. Subsequently, when she became confined to a wheelchair, she told the same surgeon that her life was still valuable to her, but if she was to become bed-ridden or incontinent it would lose its meaning and she would then prefer euthanasia. When she became incontinent and bed-ridden she vehemently stated that life still held meaning for her”. These authors suggest that this woman accommodated to the trajectory of her illness by changing her internal standards of measurement (re-calibration); she reassessed her values; and, she redefined her life quality (re-conceptualization).

This scenario depicts a patient with a deteriorating health status. Her response shift is a seemingly important coping strategy for her illness and is an adaptive mechanism.[52] It is not uncommon for these same patients with a deteriorating health status to rate their pre-treatment condition retrospectively such that they underestimate how bad they were before.[25] In some conditions, the ability to undergo an adaptive response shift does not occur. The absence of response shift has been shown in mental health conditions such as somatization or hypochondriasis.[53] Both conditions are defined as mental disorders characterized as person’s misinterpretation of physical symptoms without patho-physiological explanation. Rigidity of the patients’ conceptualization of their symptoms and their resistance to physicians’ efforts to
reformulate these maladaptive ideas prevent any normal adaptation that may improve their QoL. [53]

In summary, people strive to maintain a stable personal sense of well-being. This can be altered by situational change and be sensitive to external stimuli. Adaptation occurs due to these changes shifting their internal standards and hence, creating the possibility of response shift.[51]

5.3 Adaptation Studies on Response Shift

The existence of response shift has been documented most extensively in conditions of HIV and malignancy. The expected deterioration in QoL often does not occur in some cases of these serious illnesses. Response shift also was identified by Westerman in patients with inoperable non-small cell lung cancer treated with irradiation and chemotherapy.[54] QoL as measured by The European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ) found that contrary to expectation, there was no significant deterioration in the scale scores over the treatment period of six weeks. It seemed implausible that the side-effects of treatment had not affected their QoL.[55] Other studies also have reported counter-intuitive results. For example, patients with a life threatening disease or disability were found to report stable QoL and patients with a severe chronic illness reported QoL levels that were not inferior to that of patients with a less severe illness or to healthy patients. These counter-intuitive findings labeled by Breetvelt as “underreporting of problems” suggest that patients report less distress and dissatisfaction than they actually experience.[2, 56, 57]

Response shift has also been studied in other conditions including: a) hearing loss; b) denture therapy; (c) total joint arthroplasty; (d) sleep apnea; and, (e) stroke.[18, 21, 28, 58, 59] The direction of response shift in these conditions was an over-reporting of problems in their retrospective assessment (then-test) of their pre-morbid disability. Osborne noted that the presence of response shift can affect the interpretation of the routine pre-intervention/post-intervention assessment resulting in a greater treatment effect or a diminished treatment effect depending on whether the patient recalled a worse pre-intervention QoL or a better pre-
intervention QoL, respectively.[11] In the former case, where baseline disability is rated as being higher, a Type II error (i.e. false negative) may arise if there is no adjustment for the confounding effect of response shift.

Wagner found that children with diabetes demonstrate both over-estimation and under-estimation recalibration of pre-intervention QoL. These in turn influenced glycemic control (worsened glycemic control was noted in children with an underestimation response shift).[60] In a total knee arthroplasty population, an overestimation response shift was the most common direction compared to an underestimation response shift. The magnitude of the response shift however was small and did not alter the interpretation of the results.[61] These authors in a later study noted that with the use of a minimal clinically important difference as a threshold of change from pre to post surgery, a majority of patients did not have a response shift.[62]

5.4 Antecedents and Mechanisms for Response Shift

This section will elaborate on the response shift model of Sprangers and Schwartz that was introduced earlier.[8] Lazarus and Folkman (1984) have defined coping as “constantly changing cognitive and behavioral efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person”.[63] They defined problem-focused coping and emotion-focused coping. Emotion-focused coping strategies tend to be used when a stressor is appraised as being beyond control. These authors suggest several factors determine the coping strategy: cognitive appraisal and an individual's coping resources. The latter factor involves beliefs, cultural values, social support networks and material resources. [63] Coping strategies can contribute to the dynamism of the QoL construct. Sprangers and Schwartz suggest that an individual’s dispositional characteristics also can act as antecedents to allow for response shift. These factors can include sociodemographic characteristics such as education level, expectations and self-esteem.[4]

Expectations and satisfaction following surgery have been shown to be related to co-morbid conditions as well as patient preoperative level of physical and mental health.[62, 64] As these same variables can act as antecedents for response shift, it is possible that expectations bear a
close relationship to response shift. The Appraisal Profile of Rapkin may help unravel the cognitive processes that underlie a response shift.[7]

6.0 Degenerative Spinal Disorders

Relief of pain and disability allowing for improvement in quality of life is the primary reason for undergoing elective surgery for degenerative spinal disorders. Outcome however may be variable and, as described below, may be subject to response shift such that interpretation of effectiveness is challenging with use of conventional pre- and post-test evaluations. This section begins by describing degenerative spinal conditions, the magnitude of the problem, clinical manifestations and treatment and then describes why response shift may occur in this patient group.

6.1 Scope of the Problem and Treatment Overview

Degeneration of the spine results in disability due to both back and leg pain. The economic effect of lumbar spine disorders is immense. Lumbar spine disorders rank fifth among disease categories in the cost of hospital care and account for higher costs resulting from absenteeism from work and disability than any other category.[65] Spinal stenosis is an example of a degenerative spinal disorder; it is age dependant and its prevalence is 60% in persons age 60, 70% in persons age 70 and 80% in persons age 80.[66] In the 30 to 50 year old population, disc herniation and mechanical back pain are more common. The estimated annual incidence of sciatica in western countries is 5 cases per 1000 adults.[67] The vast majority of sciatica (over 75%) resolves with conservative treatment in patients within 3 months.[68]

When the condition is severe, the person is unable to continue with his/her vocation and/or to partake in his/her recreational pursuits. Surgery for spinal stenosis and sciatica is indicated when conservative treatment has failed. Decompression of the spinal canal by way of discectomy or laminectomy with or without fusion is the mainstay of treatment. The Spine Outcome Research Trial (SPORT), demonstrated that surgical treatment for herniated disc had significantly greater improvement in the primary outcome measures SF-36 bodily pain and SF-36-physical function
domains compared to conservative treatment at 4 years post operatively.[69] Improvement in self-reported quality of life after surgery for single level spinal stenosis is comparable to that of total hip and knee arthroplasty.[70]

6.2 Spinal Surgery and Response Shift

Outcome assessment of a treatment intervention is dependent on a valid and reliable outcome measure. Traditionally, both disease-specific and generic QoL/disability are assessed.[71] Pre-operative disability and post-operative outcomes of QoL and disability are impacted by subjective complaints (symptoms), physical findings (signs), psychological concerns and social/environmental factors.[72, 73] With the provision of spinal surgery as a catalyst and the above factors which may act as antecedents as described in the model for response shift by Sprangers and Schwartz [8], response shift possibly can occur and affect our ability to measure and interpret the outcomes following spinal surgery. To date, nothing has been published about response shift in degenerative spinal disorders. Whether response shift exists in this context and its influence on measuring outcomes is unknown.

6.3 Functional Outcome Measures in Surgery

For many years, clinicians relied exclusively on objective physical measures of disability in order to assess orthopedic surgical outcomes. Despite patient improvement in many physical abilities as a result of a surgical procedure, these measures did not adequately assess changes in functional outcomes.[74-79] Since the clinician-based tools failed to measure factors that were important to patients, functional outcome and QoL outcome measures became the foundation of measuring treatment effect. The major literature in support of subjective measures has been published during the last three decades.[80-93] Current outcome assessments for a given surgical intervention need to take into account the change in generic health status (physical, psychological, social and functional) as well as the disease-specific health status.[81, 93, 94]
6.3.1 The International Classification of Functioning, Disability, and Health

Central to this thesis is the measurement of outcomes in patients following spinal surgery. The International Classification of Functioning, Disability, and Health (ICF), provides a framework for documenting the manifestations of health conditions that result from the complex interactions of the person with the physical, social and psychological environment. The domains contained within the ICF were described from the perspective of body systems, the individual and society. According to the International Classification of Function, Disability and Health, disability is a state resulting from activity limitations and participation influenced by personal factors and one’s environment.[95]

There is no single measure that comprehensively covers all aspects of a patient’s functioning, disability, and health as represented in the ICF framework. A combination of measures that are condition-specific, generic health measures, and other dimension-specific measures often are applied in clinical studies of degenerative spinal disorders.[94]

6.4 Disease-Specific Functional Outcome Measures

The patients’ perspectives are essential in making medical decisions and judging the results of treatment. The hallmark of a disease-specific measure is the attribution of symptoms and functional limitations to a specific disease or condition.[96] Several questionnaires are available for assessing functional outcomes related to back pain. A commonly used measure recommended for use in people with spine problems[71] and used for the purpose of this thesis is the Oswestry Disability Index.[81, 93, 94]

6.4.1 The Oswestry Disability Index

The Oswestry Disability Index (ODI) is one of the most widely used outcome measures for patients with spinal disorders.[96] It consists of 10 items assessing the level of pain and interference with several physical activities, sleeping, self-care, sex life, social life, and traveling (Appendix 1). The patient selects the one statement in each section that most accurately
describes the effect of his/her pain. Each section is scored on a zero to five-point scale, with higher values representing greater disability. The sum of the ten scores (maximum score of fifty) is expressed as a percentage with higher scores representing greater disability.

Brockow showed that most concepts within the ODI could be linked to the ICF. Components of the questionnaire that represent changing basic body position link to the activities and participation component. The most frequently endorsed category of the ODI was “sensation of pain”, which links to the component of body functions.[97] The ODI also links to components of interpersonal interactions and relationships, work and employment community, social and civic life. Five concepts have been assigned to the “other specified” categories as the concepts are contained in the ICF but are not explicitly named.[98]

6.4.2 Oswestry Disability Index Measurement Properties

The ODI has demonstrated reliability and validity and compared to other back specific questionnaires, the floor and ceiling effect of the ODI is regarded as being superior.[99]

6.4.2.1 Face and Content Validity

The wording of the ODI was designed on the basis of patients’ self-reports and symptoms of chronic low back pain.[100] The content appears to adequately cover all the relevant domains. Fairbank demonstrated the validity of the ODI by the improvement in scores in twenty-five patients having their first episode of low back pain who were expected to improve with time and were shown to do so.[83] Beurskens analyzed 81 patients again confirming an expected improvement in ODI scores with improvement with back pain.[101]

6.4.2.2 Construct Validity

The ODI had a moderate correlation with pain measures such as a visual analogue scale $r= 0.62$ and the McGill Pain Questionnaire.[102, 103] The ODI has been used to validate other back pain outcome scales such as the low back outcome score.[104]
6.4.2.3 **Internal Consistency**

Fisher and Johnson found Cronbach’s $\alpha$ to be 0.76, [105] while Kopec, 1996 found this be 0.87. [105, 106] These studies show an acceptable degree of internal consistency.[100]

6.4.2.4 **Reliability**

In the original study by Fairbanks, patients were tested twice at a 24-hour interval showing a high level of agreement in scores ($n=22$, $r= 0.99$).[83] If the test-retest interval is extended to 4 days, the correlation of scores decreased to 0.91, and if retested after a week, the association was still high but decreased to 0.83.[102, 106]

6.4.2.5 **Responsiveness**

It is important that an outcome measure being used to evaluate interventions is able to detect change in response to the intervention and if the trajectory of recovery is evaluated, the outcome measure should likewise change over time. This is an instrument’s sensitivity to change or its responsiveness. The receiver-operating characteristic (ROC), tests the ability of an instrument to detect change. Sensitivity is plotted against 1 minus specificity.[107] The ROC index for the ODI was found to be 0.76. This is acceptable but less than other low back outcome instruments such as the Roland –Morris Disability Questionnaire.[100]

6.4.2.6 **Minimal Clinically Important Change (MCID)**

A statistically significant change in a score does not necessarily mean that the change is clinically important. The statistical significance of a change is sample-dependant, i.e. in large studies minute and clinically unimportant changes may be statistically significant but clinically insignificant, while the opposite may be true in small studies. In a clinical perspective, we need to evaluate whether a treatment effect is worthwhile or important when making treatment decisions.
The concept of the minimal clinically important difference (MCID) defines the smallest meaningful or important change in a clinical state.[108] There is no consensus on how much change from baseline is required before there is a clinically important change. The meaning of change has been shown to vary according to baseline entry scores and possibly other baseline characteristics.[109] For the ODI, values have ranged from 4-15 points or a 30% change from baseline.[110, 111] Ostelo et al. following an expert panel review for ranges for MCID suggested 10-12 points as the MCID.[112]

6.5 **Generic Outcome Measures in Spinal Disorders**

Generic health outcome measures broadly assess the concepts of health, disability and QoL.[113] They are used for broad comparisons of the relative impact of different conditions or treatments in the health of the population and generally, are considered to be less responsive to change than disease-specific instruments.[113] Generic measures are important because co-morbidities influence the patient’s response to treatment.[114] Generic measures also provide a more comprehensive picture of the patients’ health status because back-specific instruments do not include measures of patients’ mental or social health.

6.5.1 **Short Form 36**

The Short Form-36, (SF-36) is the most commonly used generic health status measure. It consists of 36 items that assess eight health-related concepts: physical functioning, role limitations due to physical health problems (role physical), bodily pain, general health, energy levels/fatigue (vitality), social functioning, role limitations due to emotional problems (role emotional) and psychological distress (mental health) (Appendix 2). These can be aggregated into two summary component scores: physical (SF-36-PCS) and mental health (SF-36-MCS). The SF-36 has the advantage of normative-based scoring based on very large and diverse population including individuals with back pain.[114]
6.5.1.1 **Reliability**

More than 25 studies have evaluated the test-retest reliability of the SF-36 and most have exceeded 0.80.[115-117] Reliability estimates for physical and mental summary scores are 0.92 and 0.88 respectively.[117, 118]

6.5.1.2 **Validity**

Face, content and construct (convergent and discriminate) validity have been well studied for the SF-36. Studies of validity generally support the intended meaning of high and low SF-36 scores. The SF-36 has been widely used across a variety of disease groups and studies to date have yielded content, concurrent, criterion, construct, and predictive evidence of validity.[119, 120]

6.5.1.3 **Minimal Clinically Important Difference**

The MCID for the SF-36-PCS and SF-36-MCS subscales is a change of 5 points.[116, 121]

6.6 **Life Satisfaction Questionnaire (LISAT-11)**

The Life Satisfaction Questionnaire (LISAT-11) is a checklist assessing non-health related QoL, or satisfaction and has been used by researchers and clinicians who are interested in issues of life satisfaction.[122] Items (domains) include satisfaction with: life as a whole, vocational, financial and leisure situations, contacts with friends, sexual life, self-care management, family life and partner relationships, physical health and psychological health. Each item (domain) is scored on a 6-point scale from 1, very dissatisfied, to 6, very satisfied. This scale can be dichotomized into satisfied (grades 5-6) and not satisfied (grades 1-4) and individual items measured, or a total score can be derived.[122] The scale has been shown to have test-retest reliability of 0.85-0.90.[123]
7.0 Measuring Spine Surgery Outcomes in the Face of Response Shift

The coexistence of physical, psychological and social factors is relevant to patient assessment. Perceived disability is affected by personal issues such as work demands, coping abilities and motivation.[124] At any specific time, patient assessment which is based on self-report (either preoperative or postoperative) should ideally factor in all concurrent variables that may influence his/her self-evaluation. Functional outcomes and response shift may be influenced by the presence of these psychosocial variables and other co-morbid conditions.

7.1 Impact of Co-morbid and Pre-morbid Characteristics on the Change in SF-36 and ODI

Slover demonstrated the significant impact of co-morbidities, both medical and psychosocial, on general and disease-specific health status measures following spine surgery.[125] Their results illustrated that the number and type of co-morbidities impacted the change in both SF-36 and ODI scores after lumbar spine surgery. The average change in bodily pain, physical function, physical component summary scores of the SF-36, as well as ODI scores, decreased in response to surgery as the number of co-morbidities increased. Psychosocial factors such as an active compensation case, self-rated poor health and smoking had negative effects on the change in scores after surgery. Medical disorders such as headaches, depression and nervous system disorders also were highly influential.[125] This study illustrates the difficulty in isolating true change and the importance of identifying potential confounders that can affect this.

In summary, there are number of behavioral and psychological factors that contribute to an individual’s system of dynamic standards. As clinicians, we measure health status, we monitor changes in health status and evaluate our treatment effects on health status. We use reliable and valid outcomes to measure changes in health status. However, if the target construct is shifting by way of the patients’ subjective interpretation of their health status or by way of recalibration (i.e., change in the respondent’s internal standards of measurement); reprioritization; (i.e., change in the respondent’s values); or, re-conceptualization (i.e., re-definition of the target construct), then the validity of outcome measures may be questioned. Medical and psychosocial
co-morbidities are potential causal variables or antecedents that can result in a response shift and consequently affect our ability to measure true change.

8.0 Overall Objective of the Thesis

Response shift research is lacking in outcome assessment following orthopedic interventions. The overall objective of this thesis is to determine if response shift is present in patients following decompression surgery of the lumbar spine and whether this impacts interpretation of the benefit of treatment.

The specific objectives are:

1) To determine if response shift is present in the disease specific and generic outcome measures in patients undergoing surgery of the lumbar spine.
2) To determine if response shift influences the determination of success or failure of surgery based on the minimal clinically important difference.

Secondary objectives are:

1) To determine if response shift change is consistent over time.
2) To determine if there are patient variables which may be predictors for response shift.

8.1 Hypotheses

Patients undergoing decompression lumbar surgery will undergo a response shift after surgery. Response shift will be measured as a retrospective higher estimation of their preoperative disability or as a retrospective lower estimation of their preoperative disability as determined by disease specific and generic health outcome measures. The direction of the response shift will be consistent from 6 weeks to 3 months.

Having presented the background and rationale for the thesis, the following chapter presents the manuscript describing the research. The final chapter provides a discussion of the overall thesis.
CHAPTER 2

Outcome Following Spinal Surgery: The Influence of Response Shift

1.0 Abstract

**Background:** Response shift refers to health-related changes in the self-evaluation of a concept (e.g., health, quality of life, pain) over time due to: 1) changes in internal standards (i.e., recalibration); 2) changes in values (i.e., reprioritization); or, 3) changes in conceptualization. A response shift will cause the patient to rate the same self-report measure differently pre- and post-treatment. This can result in a “moving goal post” when a self-report outcome measure is used for pre-post intervention comparisons. A “response shift” may obfuscate relevant changes of interest to clinicians. **Purpose:** The primary objective of this work was to determine if response shift is present in patients undergoing spinal surgery and, if so, to determine if it influenced the proportion of people who benefited from surgery based on achieving a minimal clinically important difference in outcome measures. Secondary objectives were to determine if there was consistency of the response shift change between 6 weeks and 3 months post surgery and to evaluate potential predictors of response shift. **Study Design/Setting:** Prospective longitudinal outcome study of patients undergoing decompression spinal surgery. **Methods:** The Oswestry Disability Index (ODI) and Short Form-36 (SF-36) were measured at baseline, 6 weeks and 3 months after spinal decompression surgery by the conventional pre-test (baseline) to post-test method. At follow-up, the then-test or retrospective pre-test score was also determined. Response shift is the difference between the pre-test score and retrospective pre-test score. Response shift was categorized into one of three categories: a retrospective overestimation of preoperative disability, a retrospective underestimation of preoperative disability, and no response shift. The minimal clinically important differences for the ODI and the SF-36 were used to define the threshold for success. The Spearman Brown coefficient for agreement was used to measure consistency of response shift between 6 weeks and 3 months. **Results:** The study cohort included 105 patients; mean age 50.6 years; 60 percent were male. At both 6 weeks and at 3 months, 72% of patients exhibited a response shift on the ODI. Overestimation of
preoperative disability was the most common direction of response shift. At 3 months, 56% and 21% of patients exhibited a response shift with the SF-36 physical (PCS) and mental (MCS) component scores, respectively. Using the ODI outcome measure, the success rate of the surgery at 3 months, as determined by the minimal clinically important difference, increased by 20 percent when response shift was taken into account compared to when it was not. Agreement between the response shift at 6 weeks and three months for the ODI and SF-36-PCS was 0.43 and 0.42, respectively, representing moderate agreement. The standard error of measurement using the SF-36-MCS was too large to make any meaningful conclusions for this outcome measure. **Conclusions:** The response shift phenomenon was shown to affect the measurement properties of standard spinal surgery outcome measures. Response shift is moderately stable from 6 weeks to 3 months. In longitudinal outcomes research following spinal decompression surgical interventions, it is critical to recognize the potential effect of response shift on the interpretation of change and on the benefit of treatment.

### 2.0 Introduction

Response shift is a psychological process whereby over time, a patient will alter his/her self-evaluation of health-related constructs such as quality of life (QoL), disability or pain. This can be by: 1) changes in internal standards (i.e., recalibration); 2) changes in values (i.e., reprioritization); or, 3) redefinition of the concept (reconceptualization).[13] A response shift will cause the patient to rate the same self-report measure differently pre- and post-treatment. Interest in response shift began in the 1990’s when clinicians began to recognize that this phenomenon could obfuscate important treatment-related changes. Response shift has been studied and recognized in patients with multiple sclerosis[126], cancer [19, 127-129], stroke [21, 58], diabetes[130, 131], dental disorders [28], and in the fields of geriatric medicine [132-134], palliative care[135-138], and orthopaedics[61]. Oort et al. demonstrated that for clinical interventions in cancer patients, adjusting for response shift increased effect sizes from moderate to large.[29]

Response shift can have implications on recovery over a period of time and can affect our measurement of disability and QoL. This has implications for measurement of cost-effectiveness,
efficacy of treatment and accuracy of the outcome measurement.[13] In parallel with reliability and validity studies that examine measurement properties of the disease-specific and general-health outcome measure, evaluation of response shift is needed to capture the dynamic change in the illness experience and perspectives of the patients.[139]

The dispositional and disease factors that may impact on response shift are not well defined. The model of Sprangers and Schwartz proposes that stable personality characteristics (i.e., antecedents) interact with cognitive, affective, or behavioral processes that individuals use to deal with life changes (i.e., mechanisms) and promote response shifts.[8] This results in a level of perceived QoL change that is higher or lower than expected based upon objective criteria. This process can occur with changes in health status associated with progression of a chronic condition, an acute change in health, or as a result of treatment interventions (catalyst). The response shift process is hypothesized to occur continuously over time as part of a positive or negative adaptation process.[14]

Spine outcome research has benefited from uniform standards for measuring patient-reported outcomes as recommended by Deyo.[71] These include the Short Form-36 (SF-36)[140], numeric rating scales for pain such as visual analogue scales (VAS), and disease-specific functional outcome measures such as the Oswestry Disability Index (ODI).[141] Response shift may be a form of bias if a patient has changed the way s/he is using the measurement scale from one time point to the next and hence be a threat to the internal validity of the outcome questionnaire when used for longitudinal prospective studies.

Most clinicians anecdotally report observing response shift. For example, it is not uncommon for spine surgeons to observe that, although their patients express satisfaction with the surgery and would do it again with 20:20 hindsight, their back pain, leg pain or function is not scored as well as would be expected on outcomes such as a pain VAS or the ODI.[73] This type of inconsistency may reflect changes in values or conceptualization of QoL. This type of contradiction was previously reported by Schwartz in young adult survivors of cancer whose QoL reportedly declined after a psychosocial intervention gave them more confidence to engage in physically challenging activities.[12] A closer look at the results, however, indicated that the
young adult survivors of cancer shifted in how physical functioning was valued and this reappraisal of their values contributed to their global reassessment of QoL. These patients now placed more value on the physical functions that they could not do before and they valued these physical parameters in a way more similar to healthy individuals. Thus the intervention was indeed successful because the cancer survivors were now functioning more similarly to healthy peers, but when reexamining their QoL pre-treatment, they now rated it worse.[142]

In the context of the spine surgery patient, an operation may be a success because s/he can partake in physical activity that they previously could not. However, if the same adaptation occurs as in the preceding scenario, when looking back at their QoL prior to their surgery, this new ability may result in the individual recognizing how much they could not do before. Hence, retrospectively, they may assess their previous QoL as being worse than they realized preoperatively. This is the essence of response shift and can result in a patient interpreting and responding to the same questionnaire in a different way over time when there has been a change in their health status.

The purpose of this study was to determine if the response shift phenomena is present in patients undergoing spinal surgery. If response shift is present, its effect on the interpretation of outcome was classified with respect to the minimal clinically important differences (MCID) in the ODI and SF-36.[109, 110] Secondary objectives were to determine if there is consistency in the response shift at two time points following surgery and what patient factors may be predictive of response shift. The primary outcome measure was the back disease-specific ODI. Secondary outcome measures were the SF-36 physical component score (PCS) and SF-36 mental component score (MCS).

3.0 Methods

3.1 Study Design and Setting

The study is a longitudinal observational cohort study. Patients were recruited from the clinics of three spine surgeons in the Division of Orthopaedics at Sunnybrook Health Sciences Center, a
tertiary care hospital associated with the University of Toronto. All surgeries were performed by one of these three surgeons. The enrollment period for the study cohort was from January 2007 to March 2008.

3.2 Participants

Patients undergoing elective posterior lumbar spinal decompression surgery for either one or two level spinal stenosis or for disc herniation were candidates for recruitment. The clinical indication for surgery was leg dominant pain without lumbar instability. Concurrent back and leg pain was common, however, back dominant pain was an exclusion for decompression surgery. Patients requiring fusion, which is a common treatment for back dominant pain, were excluded.

All patients were adult (over 18 years of age). There was no upper age exclusion criterion. Patients were excluded from the study if they were unable to complete the questionnaires in English, if they had visual or cognitive impairment or any other disability that prevented them from completing the outcome measures independently, or if they were unable to give consent for participation. Eligible patients were approached by research personnel not involved in their care and informed consent was obtained. This study was approved by REB at Sunnybrook Health Sciences Center. (A copy of the REB form is contained in Appendix 3.)

3.3 Data Collected

3.3.1 Primary and Secondary Outcomes

The primary outcome for this study was the Oswestry Disability Index (ODI). The ODI is a disease-specific measure that evaluates limitations in various activities of daily living for patients with low back pain. The ODI is a 10-item questionnaire, with each item scored on a “0-5” scale. The ten items are summated (maximum total of 50) and converted to a percentage. The higher the score, the greater the disability.[83, 94] The secondary outcome measures used in this study were the SF-36-PCS and SF-36-MCS. The SF-36 was used to measure changes in generic health status before and after surgery.[93, 118] The outcome measures used in this study are those
traditionally collected in spinal surgery[71] and have demonstrated reliability and validity.[81, 93, 94, 102, 103, 105, 106]

3.3.2 Descriptive Data

Further data collected were patient demographics (age and sex), duration of symptoms, employment status (working at present, retired, student, homemaker), and compensation status (currently on disability or compensation). Associated co-morbid health conditions as well as other musculoskeletal conditions were collected from the patient’s past medical history during the initial clinical interview.

3.3.3 Potential Predictor Variables of Response Shift

Variables were assessed as being potential predictors of response shift. These were the life satisfaction questionnaire (LISAT -11), pre-operative VAS leg pain score, pre-operative VAS back pain score, duration of symptoms, age, sex, compensation status, number of co morbidities and pre-operative ODI score.

LISAT-11 is a checklist assessing non- health-related QoL or satisfaction with items: life as a whole, vocational, and financial and leisure situations, contacts with friends, sexual life, self-care management, family life and partner relationships, physical health and psychological health. (Appendix 4).[122, 123] Each item is scored on a 6-point scale from 1 very dissatisfied, to 6, very satisfied. A total summed score is derived and is a measure of life satisfaction.[122] The scale has been shown to have test-retest reliability of 0.85- 0.90.[123] Visual analogue scales (VAS) for both back pain and leg pain were scored on a linear line between 0 and 100. This score was used to determine the dominant symptom and quantify the pain perceived by the patient.
4.0 Procedures

Participants completed the questionnaires pre-operatively, at 6 weeks post-operatively and at three months post-operatively. Pre-operative questionnaires were completed in the clinic within two weeks prior to surgery and took approximately 20-30 minutes to complete. After completion, questionnaire booklets were checked for completeness prior to the patient leaving the clinic and the study assistant gathered any missing information.

Participants returning for routine follow-up also completed questionnaires at designated follow-up times. However, due to varying follow up protocols of the surgeons, participants who did not return to the Orthopaedic Surgery clinic at the prescribed follow-up time points were mailed the questionnaires. Any questionnaires that were found to be incomplete were flagged and the patients were subsequently called to obtain the missing results.

Pre-operatively, participants completed the questionnaire package containing the ODI and SF-36 based on their current status. Post-operatively, they completed the same questionnaires at each follow-up assessment. Follow-ups were carried out at 6 weeks and at 3 months. Three months is a clinically meaningful time following decompression surgery, as recovery and rehabilitation are generally near complete at this time. Using structured questions at follow-up, the research assistant also asked the patients for any changes in their demographic or co-morbid health information.

At each follow-up assessment, the first questionnaire asked the patient to respond according to how they currently felt (post-test). The second questionnaire for measuring response shift was the then-test, completed at the same time as the post-test. The then-test was used as the method for measuring RS as it is easy to use, it is applicable to individual patients, the analysis does not require a large sample size and it is the only method which allows quantification of response shift. The then-test requires the patient to rate the pain and disability that they had prior to their surgical treatment. Instructions for the then-test were: “please complete the questionnaire with respect to how you felt prior to your surgery”. In other words, the patient was asked to provide a renewed judgment of their condition before surgery. Both packages (post-test and the then-test)
contained the same outcome measures as the pre-test (ODI, SF-36). The post-test and then-tests were done at the same time point, which presumes that patients used the same internal scale of measurement when answering both sets of questionnaires.[14] Additionally, it is assumed that the patient will not have remembered the answers they gave on the pre-test, but that they recall their preoperative level of disability and QoL.[4, 15]

Data entered into the database were checked with the hard copy on a random sample of fifteen patient files. This was performed every four months and the results of the outcome scores were confirmed through manual calculations. A Microsoft Access database was used for the entry of all collected data. A unique identification number identified all participants. All data was stored on a password-protected database.

5.0 Analysis

Descriptive statistics were calculated for all variables of interest. Continuous measures were summarized using means and standard deviations whereas categorical measures were summarized using counts and percentages. Univariable tests, t-test and chi-square as appropriate to the type of data, were used for comparisons between the excluded to included participants, and for comparisons between characteristics of patients with one vs. two follow-up assessments.

The then-test was used for measurement of response shift. The difference between the initial pre-test values and the retrospective pre-test values is the amount of response shift. A patient's retrospective re-evaluation of their baseline status can be the same, worse or better than their prospective baseline assessment.[11, 143, 144]

The conventional treatment effect (TE) is measured as the difference between the post-op rating and the pre-op rating prior to the surgery. This is the routine method of measuring change following treatment. The adjusted treatment effect (ATE) is measured as the difference between the post-op rating and the then-test rating. The equations are summarized in Table 1.
Table 1. Summary Equations

<table>
<thead>
<tr>
<th>Response Shift</th>
<th>Pre-test – then-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted Treatment Effect</td>
<td>Post-test – then-test</td>
</tr>
<tr>
<td>Unadjusted Treatment Effect</td>
<td>Post-test – pre-test</td>
</tr>
</tbody>
</table>

Response shift scores (pre-test – then-test scores) that were +/- 2.5 SD from the mean of the sample for each outcome measure were considered outliers and their results were potentially invalid for reasons such as misunderstanding the directions. These scores were excluded from the analysis.

5.1 Minimum Detectable Change

Sensitivity of the outcome measures to change was evaluated using the Minimal Detectable Change (MDC). MDC is a change that is greater than the standard error (SEM) of the measurement instrument. Individual level MDC provides information on whether observed changes in the ODI, or SF-36-PCS or SF-36-MCS are greater than chance variations.[145] A change of less than the MDC for any measure was considered as no change. It is calculated based on the standard deviation (SD) of the cohort being studied and the test-retest reliability coefficients of the outcome instrument (r).[146, 147]

\[
\text{MDC} = \text{SEM} \times 1.96 \times \sqrt{2} \text{ (95% CI)} ; \quad \text{SEM} = SD \times \sqrt{1 - r}.
\]

The MDC used for this cohort was calculated based on the standard deviation of the pre-operative raw scores of the ODI, SF-36-PCS and SF-36-MCS scores. The MDC, the standard deviations and the test-retest reliability coefficients, used for the outcome measures are summarized in Table 2.[83, 96, 102, 105, 115, 117] MDC reflects statistically significant change
and is not equivalent to a clinically important change, which is discussed below. A change in pre-test score to then-test score for a given individual that was below the MDC was considered as having no response shift.

**Table 2: Minimal Detectable Changes, Standard Deviations and Test-Retest Reliability Coefficients of Outcome Measures**

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Test-retest (r) reliability coefficient</th>
<th>Standard Deviation</th>
<th>Minimum Detectable Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODI</td>
<td>.94</td>
<td>7.2</td>
<td>5</td>
</tr>
<tr>
<td>SF-36-PCS</td>
<td>.90</td>
<td>10.5</td>
<td>8</td>
</tr>
<tr>
<td>SF-36-MCS</td>
<td>.88</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

MDC was used to categorize change in participants’ scores and was the basis of response shift classification into the following groups

1) No RS (difference < MDC).
2) Retrospective overestimation of baseline disability. (Using ODI, then-test is greater than pre test. For SF36-PCS and MCS, then-test is less than pre-test scores).
3) Retrospective underestimation of baseline disability. (For ODI, then-test less than pre-test. For SF-36-PCS and MCS, then-test is greater than pre-test scores).

### 5.2 Minimal Clinically Important Difference

Standards for minimal clinically important difference (MCID) are required for interpretation of intra-individual change as having clinical meaning in the observed change in individual scores.[121] MCID is the minimal clinically detectable change that is considered clinically significant.[148] A measurement scale can only be sufficiently sensitive to detect meaningful change if the values of MDC do not exceed the values of MCID. For the ODI, there is no consensus on how much change from baseline is required before there is a clinically important
difference. Values have ranged from 4-15 points or a 30% change from baseline. Ostelo, following an expert panel review for ranges for MCIDs for the ODI, suggested 10-12 points as the MCID.[112] Hagg et al. concluded that the MCID for ODI was 10 points [110]. For purposes of this study a 10 point change in ODI scores was considered a MCID. However, sensitivity analysis was also conducted using a 12- point change in ODI as a threshold value for MCID. The MCID for SF-36-MCS and SF-36 PCS scores was a change of 5 points.[116, 121] The sensitivity of the SF-36- PCS and SF-36- MCS scales to individual change was very low, as indicated by the higher values of MDC, 8 and 14 respectively shown in Table 2. Consequently, I could not use the MCID for the SF-36 scales.

5.3 Consistency of Response Shift

Those patients who had two response shift measures, at times 6 weeks and at 3 months were used for this analysis. Consistency or stability of response shift at the two time periods was measured by the Spearman-Brown coefficient as the two measures of response shift are not independent (same patient at two time points). With lack of independence apparent agreement may reflect more a recollection of the previous decision than a genuine renewed judgment as is required for the then-test. An artificially high kappa coefficient can be obtained if there is a lack of independence. Agreement of response shift direction was determined (overestimation, underestimation or no response shift) from 6 weeks to three months. For exploratory research, 0.60 is commonly used as a measure of good reliability for the Spearman-Brown Coefficient. The interpretation of agreement was based on established Kappa values (0.01-0.20 is slight, 0.21-0.40 is fair, 0.41- 0.60 is moderate, 0.61-0.80 is substantial and 0.81-1.00 is high).[149-151]

5.4 Predictor Variables

The means for each of the directions of response shift, (overestimation of preoperative disability, underestimation of pre-operative disability and no response shift) was compared to determine if any significant differences were present between these based on the potential predictor variables.
Analysis of variance was used for continuous variables and Chi-square for discrete variables. No multivariate analysis could be performed due to the small sample size.

5.5 **Effect Size**

Magnitude of change in ODI and SF-36-PCS and SF-36-MCS scores was assessed using Cohen’s $d$. [152] Cohen’s $d$ is a standardized measure of effect size (ES) and provides information on the amount of change in the measure relative to the variation with the measure. For paired data, Cohen’s $d$ is computed as the difference between the baseline and the follow-up scores divided by the standard deviation of the difference in the scores. ES values of 0.20-0.49 are considered small, 0.50-0.79 are moderate and $\geq$ 0.80 are large.[152]

The ES was calculated based on the conventional treatment effect (pre- to post-scores) and the adjusted treatment effect scores (then- to post-scores). Statistical difference between the unadjusted and adjusted treatment effect was tested using a paired t-test.

5.6 **Sample Size Calculation**

There was no preliminary data on which to base sample size calculations. A post-hoc power calculation was performed.

All analyses were conducted using the Statistical Package for the Social Sciences (SPSS) version 16.0, 2007 software (SPSS® Chicago, IL).

6.0 **Results**

6.1 **Patient Accrual**

Figure 3 describes patient accrual. One hundred and seventy eight patients were approached and asked to consider participation. Thirteen patients were excluded due to lack of English fluency.
Of the 165 eligible patients, 11 failed to return the pre-operative questionnaires, 3 withdrew from the study, and 38 failed to attend follow-up appointments and did not return questionnaire sent via mail. Eight patients completed the follow-up post-test successfully, but failed to complete the then-test package. One hundred and five patients had complete data for either the 6-week (82 patients) or 3-month (86 patients) time period. Sixty-three patients had complete data for both 6-week and 3-month assessments.

**Figure 3: Patient Accrual**

6.2 **Out of Study Eligible Patients Data**

For the out of study patients, patient-characteristic data were available on 46 of 60 patients. These were compared with those of the in-study patient group to determine if there was any evident selection bias. These two groups had similar levels of performance on all of the outcome variables at baseline. There were no statistically significant differences in age, number of co-morbidities, pre-operative ODI, pre-operative SF-36 -PCS, pre-operative SF-36-MCS or duration of symptoms (Table 3). There was no statistically significant difference in sex, however, there were more males in the included group and the lack of any statistically significant difference
may be due to the small sample size. There was also no difference in percentage of patients on compensation, 25% in each group. The most common co-morbidity for both groups was lower extremity musculoskeletal conditions. The similarities between the in-study and out of study patient groups suggest there was no bias in the included sample cohort.

**Table 3: Comparative Data for the In-Study vs. Out of Study Patients**

<table>
<thead>
<tr>
<th></th>
<th>In-study (N=105)</th>
<th>Out of Study (N=462)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age<strong>1</strong></td>
<td>50.6</td>
<td>50.0</td>
<td>.82</td>
</tr>
<tr>
<td>% Male<strong>2</strong></td>
<td>60.0</td>
<td>45.7</td>
<td>.57<strong>3</strong></td>
</tr>
<tr>
<td>Pre ODI</td>
<td>47.7</td>
<td>48.4</td>
<td>.88</td>
</tr>
<tr>
<td>Pre SF36-PCS<strong>1</strong></td>
<td>26.0</td>
<td>23.4</td>
<td>.84</td>
</tr>
<tr>
<td>Pre SF-36-MCS<strong>1</strong></td>
<td>39.8</td>
<td>42.2</td>
<td>.60</td>
</tr>
<tr>
<td>Pre VAS back<strong>1</strong></td>
<td>53.8</td>
<td>56.3</td>
<td>.56</td>
</tr>
<tr>
<td>Pre VAS leg<strong>1</strong></td>
<td>63.6</td>
<td>68.6</td>
<td>.46</td>
</tr>
<tr>
<td>Duration of Symptoms (months)</td>
<td>21.8</td>
<td>19.7</td>
<td>.84</td>
</tr>
</tbody>
</table>

**1**Means

**2**46 / 60 of the out of study patients had demographic data for comparison to the included cohort

**3**Chi-square statistic .32 (df-2); NS

Due to different surgeon practices, the patients had their initial post-operative follow-up at either 6 weeks or at 3 months. For those with a follow-up at 6 weeks, many returned also at 3 months. Table 4 compares the baseline patient variables for those with one post-operative follow-up assessment (6 weeks or 3 months; N=82) and those with two post-operative follow-up assessments (6 weeks and 3 months; N=63). There were no significant differences in any of the characteristics based on follow-up. The group with two post-operative follow-up assessments is thus representative of the entire cohort.
Table 4: Characteristics of Patients with One Vs. Two Follow-up Assessments

| Characteristic          | One Time point (n=82) | Two time points (n=63) | Statistic | P-value  
|-------------------------|-----------------------|------------------------|-----------|---------
| Age                     | 50.3                  | 50.7                   | .48       | .63     
| Sex (%male)             | 56.1                  | 62.3                   | 4.80      | .09     
| Duration of Symptoms    | 18.2                  | 23.4                   | .41       | .16     
| Pre operative ODI       | 47.5                  | 47.3                   | .33       | .74     
| Pre operative SF-36-PCS | 25.9                  | 26.3                   | .13       | .90     
| Pre operative SF-36-MCS | 41.8                  | 38.6                   | 1.00      | .31     
| Pre operative VAS - Back| 57.0                  | 51.6                   | .63       | .53     
| Pre operative VAS - Leg | 62.7                  | 64.3                   | .33       | .74     

1 t statistic is used for all comparisons except sex
2 Chi-square statistic; NS

Table 5 indicates the most common co-morbid conditions in the included sample.

Table 5: Most Common Co-morbid Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Percentage (n=105)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Limb Arthritis</td>
<td>20.2</td>
</tr>
<tr>
<td>Depression</td>
<td>7.7</td>
</tr>
<tr>
<td>Diabetes</td>
<td>6.7</td>
</tr>
<tr>
<td>Cancer</td>
<td>5.8</td>
</tr>
<tr>
<td>Heart Condition</td>
<td>5.8</td>
</tr>
</tbody>
</table>

6.3 Outcome Scores

The ODI, SF-36-PCS and SF-36-MCS measures all showed significant improvements from pre-operative to post-operative follow-up times (p<0.01). The mean then-test scores behaved similarly for all outcome measures. For the overall cohort, the average then-test scores indicate an overestimation of the retrospective rating of disability compared to the pre-operative ratings (i.e. higher retrospective scores for the ODI and lower retrospective scores for the SF-36-PCS and SF-36-MCS compared to preoperative scores). The mean preoperative, then-test and the
postoperative scores for the ODI, SF-36-PCS and SF-36-MCS are illustrated in Tables 6 and 7 for the entire cohort of patients. Response shift, (the difference between the pre-operative scores and the then-test scores) was statistically significant for the ODI and SF-36-PCS (p =0.001) but not for the SF-36-MCS (p=0.078) at both 6 weeks and at 3 months.

**Table 6: 6 Week Summary Scores**

<table>
<thead>
<tr>
<th></th>
<th>ODI</th>
<th>SF-36-PCS</th>
<th>SF-36-MCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td><strong>Pre-op</strong></td>
<td>105</td>
<td>47.7 (7.2)</td>
<td>26.0 (10.5)</td>
</tr>
<tr>
<td><strong>6 Week</strong></td>
<td>82</td>
<td>23.0 (7.5)</td>
<td>36.7 (11.4)</td>
</tr>
<tr>
<td><strong>6W then-test</strong></td>
<td>82</td>
<td>52.4 (7.6)</td>
<td>22.2 (11.1)</td>
</tr>
<tr>
<td><strong>Mean RS</strong></td>
<td>82</td>
<td>4.7 (p&lt;.001)</td>
<td>3.8 (p&lt;.001)</td>
</tr>
</tbody>
</table>

* Significant Response Shift

**Table 7: 3 Month Summary Scores**

<table>
<thead>
<tr>
<th></th>
<th>ODI</th>
<th>SF-36-PCS</th>
<th>SF-36-MCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td><strong>Pre-op</strong></td>
<td>105</td>
<td>47.7 (7.2)</td>
<td>26.0 (10.5)</td>
</tr>
<tr>
<td><strong>6 Week</strong></td>
<td>86</td>
<td>23.1 (8.0)</td>
<td>36.7 (11.4)</td>
</tr>
<tr>
<td><strong>6W then-test</strong></td>
<td>86</td>
<td>54.3 (10.1)</td>
<td>22.2 (11.1)</td>
</tr>
<tr>
<td><strong>Mean RS</strong></td>
<td>86</td>
<td>4.7 (p&lt;.001)</td>
<td>3.8 (p&lt;.001)</td>
</tr>
</tbody>
</table>

* Significant Response Shift
The effect size of the ATE is greater than for the TE for all outcome measures at both 6 weeks and at 3 months where a significant response shift was found. Table 8 summarizes the ATE and TE for each outcome measure as well as the effect sizes.

**Table 8: Adjusted and Unadjusted Treatment Effects and Effect Size**

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Mean</th>
<th>Effect Size&lt;sup&gt;4&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Week ODI ATE&lt;sup&gt;1&lt;/sup&gt;</td>
<td>-29.3&lt;sup&gt;3&lt;/sup&gt;</td>
<td>1.3</td>
</tr>
<tr>
<td>6 Week ODI TE&lt;sup&gt;2&lt;/sup&gt;</td>
<td>-24.1</td>
<td>1.2</td>
</tr>
<tr>
<td>3 Month ODI ATE</td>
<td>-31.4</td>
<td>1.3</td>
</tr>
<tr>
<td>3 Month ODI TE</td>
<td>-25.6</td>
<td>1.2</td>
</tr>
<tr>
<td>6 Week PCS ATE</td>
<td>14.5</td>
<td>1.2</td>
</tr>
<tr>
<td>6 Week PCS TE</td>
<td>10.8</td>
<td>1.1</td>
</tr>
<tr>
<td>3 Month PCS ATE</td>
<td>13.7</td>
<td>1.1</td>
</tr>
<tr>
<td>3 Month PCS TE</td>
<td>10.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<sup>1</sup> Adjusted treatment effect; <sup>2</sup> Unadjusted treatment effect
<sup>3</sup> ODI is a negative value as improvement is indicated by a lower number. For PCS and MCS, improvement is a higher score.
<sup>4</sup>ES values of 0.20-0.49 are considered small, 0.50-0.79 are moderate and ≥ 0.80 are large.<sup>[162]</sup>

### 6.4 Post-hoc Power Calculation

The effect of interest for determining response shift for the primary outcome measure is a change in ODI of 5 points (this corresponds to a change ≥MDC). The mean ODI response shift was 5.8 points when combining 6 week and 3 month data together. The power to detect a difference of 5.8 was 0.43. While the power was low, there was a significant response shift and large effect size and, hence, the low power was not an issue (i.e. power would only be important if no statistically significant change was detected as the chance of a Type II error, a false negative, would be high).

In the SF-36-PCS outcome measure, the power of the sample was 0.27. The mean response shift was found to be 2.7 and was statistically significant. While the power was low, there was a statistically significant response shift and large effect size and, similar to the ODI, the low power
was not an issue. Power was too low to exclude the possibility of a type II error in the measurement of the response shift for the SF-36-MCS given that the difference between the pre-test and then-test was not statistically significant.

6.5 **Frequency and Direction of Response Shift**

The frequency and percentage of patients undergoing a response shift for ODI and SF-36-PCS is illustrated in Tables 9 and 10. At both times, the generic PCS had more people experiencing no response shift than the disease-specific ODI. Overall, the most prevalent direction of the response shift was a retrospective overestimation of their preoperative disability.

**Table 9: Direction of Response Shift: ODI**

<table>
<thead>
<tr>
<th>Direction of Response Shift</th>
<th>Frequency /($)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>6W</strong></td>
<td><strong>3M</strong></td>
</tr>
<tr>
<td>No Response Shift$^1$</td>
<td>23 / (28.0)</td>
</tr>
<tr>
<td>Retrospective higher estimation of pre-op disability</td>
<td>43 / (52.4)</td>
</tr>
<tr>
<td>Retrospective lower estimation of pre-op disability</td>
<td>16 / (19.5)</td>
</tr>
<tr>
<td><strong>Total Number</strong></td>
<td><strong>82 / (100)</strong></td>
</tr>
</tbody>
</table>

$^1$RS < MDC
Table 10: Direction of Response Shift: SF-36-PCS

<table>
<thead>
<tr>
<th>Direction of Response Shift</th>
<th>Frequency /(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6W</td>
</tr>
<tr>
<td>No Response Shift(^1)</td>
<td>57/ (69.5)</td>
</tr>
<tr>
<td>Retrospective higher estimation of pre-op disability</td>
<td>19/ (23.2)</td>
</tr>
<tr>
<td>Retrospective lower estimation of pre-op disability</td>
<td>6/(7.3)</td>
</tr>
<tr>
<td>Total Number</td>
<td>82/ (100)</td>
</tr>
</tbody>
</table>

\(^1\) RS< MDC

In the SF-36-MCS outcome scale few people reported differences in their pre-test and then-test scores greater than MDC of 14 points. Only 17% and 22% respectively at 6 weeks and 3 months had a response shift (either higher or lower than the pre-operative score) that could be considered to be statistically significant. Table 11 illustrates the direction of response shift for this outcome measure with the overwhelming majority in the no response shift category.
<table>
<thead>
<tr>
<th>Direction of Response Shift</th>
<th>6W (%)</th>
<th>3M (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Response Shift¹</td>
<td>68/ (82.9)</td>
<td>67/ (77.9)</td>
</tr>
<tr>
<td>Retrospective higher estimation of pre-op disability</td>
<td>8/ (9.8)</td>
<td>16/(18.6)</td>
</tr>
<tr>
<td>Retrospective lower estimation of pre-op disability</td>
<td>6/(7.3)</td>
<td>3/ (3.5)</td>
</tr>
<tr>
<td>Total Number</td>
<td>82/ (100)</td>
<td>86/ (100)</td>
</tr>
</tbody>
</table>

¹ RS< MDC

6.6 Consistency of Response Shift at 6 Weeks and 3 Months

When looking at the individual patients, forty-one out of 63 patients (65.0%) had the same categorization of response shift at both time points for the ODI. The Spearman Brown Correlation for agreement or consistency was 0.43, which corresponds to moderate agreement.[149] Table 12 shows the cross tabulation for response shift categorization at the two time points for the ODI.
Table 12: ODI Consistency of Response Shift Direction at 6 Weeks and 3 Months

<table>
<thead>
<tr>
<th></th>
<th>NO RESPONSE SHIFT</th>
<th>OVER ESTIMATION</th>
<th>UNDER ESTIMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>6W RESPONSE SHIFT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO RESPONSE SHIFT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERESTIMATION</td>
<td>6</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>UNDERESTIMATION</td>
<td>5</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

1 Response shift < MDC

Thirty-nine of 63 patients (61.9%) had the equivalent response shift categorization at both time points for SF-36-PCS. The Spearman Brown Correlation for agreement was 0.42, corresponding to moderate agreement. Table 13 shows the cross tabulations for response shift category at the two time points for SF-36-PCS.

Table 13: SF-36-PCS Consistency of Response Shift Direction at 6 Weeks and 3 Months

<table>
<thead>
<tr>
<th></th>
<th>NO RESPONSE SHIFT</th>
<th>OVER ESTIMATION</th>
<th>UNDER ESTIMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>6W RESPONSE SHIFT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO RESPONSE SHIFT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERESTIMATION</td>
<td>0</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>UNDERESTIMATION</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

1 Response shift < MDC

No measurement of agreement for the SF-36-MCS measure could be performed as 77.8% and 84.1% respectively at 6 weeks and 3 months did not show any response shift due to the high MDC of 14.
6.7 **Minimal Clinically Important Difference**

Table 14 illustrates the change in interpretation of the clinical outcome using the ODI when comparing the conventional unadjusted treatment effect (pre-op to post-op) to the adjusted treatment effect (which accounts for response shift). The clinical status of a patient can be worse post operatively, unchanged or better. A higher ODI score post-operatively would measure deterioration in clinical status. Seven patients did deteriorate after surgery based on the TE and six were worse based on the ATE. The number of patients who were unchanged following surgery (change in ODI <= MDC) was altered by response shift. Similarly the number of patients who were improved, but less than MCID was affected by response shift.

The most important consideration in interpretation of clinical outcome is the proportion of patients that are improved by the MCID. The number of patients who improved above the MCID increased from 55 (66.3%) to 66 (79.5%) when accounting for response shift. Four patients were excluded from this analysis because their response shift scores were outliers, leaving 83 patients for this analysis. When the four outliers are also included in the results, the number of patients who improved above MCID increased from 58 (67.4%) to 68 (79.1%) patients. One outlier had a retrospective underestimation of preoperative disability, while two had an overestimation of preoperative disability. Inclusion of outliers had no material effect on the results.

For a sensitivity analysis of the MCID, a 12 point difference was used in addition to a 10 point difference in the ODI. Both these values have been reported in the literature as the MCID for ODI.[112] As was expected, with a higher threshold for MCID, (12 vs.10), the change in number of patients who had a clinically significant outcome based on the ATE but not on the TE reduced from 11 to 10 patients.
Table 14: Minimal Clinically Important Difference - 3 Month ODI

<table>
<thead>
<tr>
<th>Clinical Status</th>
<th>Treatment Effect</th>
<th>Adjusted Treatment Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deteriorated &gt; MDC</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>No change</td>
<td>12</td>
<td>7 (8)</td>
</tr>
<tr>
<td>Improved &lt; MCID</td>
<td>9 (10)</td>
<td>4 (5)</td>
</tr>
<tr>
<td>Improved &gt; MCID</td>
<td>55 (58)</td>
<td>66 (68)</td>
</tr>
</tbody>
</table>

1 ODI change is less than Minimal Detectable Change in either direction.
2 Pre to post change
3 Then test to post change (accounts for response shift)
4 Minimal Clinically Important Difference
5 Bracketed values include the four outliers.

6.8 Predictor Variables

The variables age, sex, number of co-morbidities and duration of symptoms for the outcome measure SF-36-PCS at 3 months were the only statistically significant predictor variables comparing the three categories of response shift in bivariate analysis. (Table 15)

Table 15: P-Values for the Comparison of Means for Potential Predictor Variables

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>6W ODI</th>
<th>3M ODI</th>
<th>6W PCS</th>
<th>3M PCS</th>
<th>6W MCS</th>
<th>3M MCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.06</td>
<td>0.75</td>
<td>0.70</td>
<td>0.01*</td>
<td>0.45</td>
<td>0.92</td>
</tr>
<tr>
<td>Gender</td>
<td>0.51</td>
<td>0.43</td>
<td>0.37</td>
<td>0.01*</td>
<td>0.75</td>
<td>0.27</td>
</tr>
<tr>
<td># Co-morbidities</td>
<td>0.95</td>
<td>0.44</td>
<td>0.69</td>
<td>0.03*</td>
<td>0.97</td>
<td>0.49</td>
</tr>
<tr>
<td>Duration of Symptoms</td>
<td>0.16</td>
<td>0.89</td>
<td>0.08</td>
<td>0.03*</td>
<td>0.43</td>
<td>0.77</td>
</tr>
<tr>
<td>Pre Back VAS</td>
<td>0.96</td>
<td>0.33</td>
<td>0.24</td>
<td>0.46</td>
<td>0.78</td>
<td>0.47</td>
</tr>
<tr>
<td>Post Back VAS</td>
<td>0.63</td>
<td>0.65</td>
<td>0.38</td>
<td>0.17</td>
<td>0.76</td>
<td>0.691</td>
</tr>
<tr>
<td>Pre ODI</td>
<td>0.66</td>
<td>0.15</td>
<td>0.87</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post ODI</td>
<td>0.09</td>
<td>0.15</td>
<td>0.27</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compensation</td>
<td>0.54</td>
<td>0.38</td>
<td>0.32</td>
<td>0.67</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>LISAT-11</td>
<td>0.98</td>
<td>0.77</td>
<td>0.74</td>
<td>0.25</td>
<td>0.43</td>
<td>0.94</td>
</tr>
</tbody>
</table>

* Significant
7.0 Discussion

7.1 Primary and Secondary Objectives

This study demonstrated that response shift is present in patients following spinal surgery, and is the first to identify the presence of this phenomenon in this patient population. The potential change in the interpretation of treatment effectiveness as a result of response shift is the most important finding from this study. A documented change in an outcome measure of interest can have no statistical significance if the change is less than MDC. If statistical significant change is found, the change must have a clinically significant effect as measured by the MCID. Response shift causes patients to change the way they assess their health or perception of disability over time. Capturing this change in conjunction with the raw measurement of outcome is critical to the interpretation of success or failure of treatment.

If response shift is a bias when using self-administered questionnaires, its identification may be considered a threat to the internal validity of any self-administered questionnaires when they are used to measure change over time. When a retrospective larger estimation of one’s previous disability is present, a traditional pre- and post-treatment comparison that does not adjust for response shift will underestimate the value of a given procedure.

The concept of the minimal clinically important difference (MCID) defines the smallest meaningful or important change in a clinical state. A MCID must occur to consider surgery a success. If a patient is not improved by this threshold based on the conventional pre- to post-treatment measurement, but at the same time undergoes a response shift with a higher retrospective estimation of preoperative disability, this will potentially lead to a Type II error, (incorrectly concluding there is no significant effect). In this study, at three months, it was found that 28 patients out of 83 did not reach the MCID for the ODI score when response shift was not measured. When response shift was accounted for, 11 more patients crossed the MCID threshold. When accounting for response shift, a greater effect size was found, as well as an overall improved outcome rate. When accounting for response shift the findings would indicate a greater cost effectiveness of treatment.
Response shift in the opposite direction (retrospective lower estimation of pre-operative disability) if not measured, can potentially overestimate the value of surgery. In this case conventional pre- to post-measurement may be suggestive of a change in functional outcome greater than MCID; however a change in patient perspective may in theory actually cause a change less than the MCID. This direction of response shift was infrequently seen in this study and is more commonly seen in patients with deteriorating health conditions such as malignancy and HIV infection. This pattern of response shift can form a valuable strategy for coping with a chronic disease by recalibrating one’s expectations for health and functioning.[11] In my study and others that similarly demonstrate an overestimation of pre treatment-disability, the clinical or psychological utility is less well understood. Regardless of the direction, measuring response shift for any change in a patient’s perspective is important to determine a non-confounded treatment effect.

There is no single instrument that comprehensively covers all aspects of a patient’s functioning and disability. A combination of measures that evaluate QoL, disability and pain are commonly applied in clinical studies of degenerative spinal disorders. Condition-specific (e.g., the ODI) and generic outcome measures (e.g., the SF-36) are used in parallel.[39, 94, 112] In my study, the number of patients demonstrating response shift was dependant on the questionnaire used. The disease-specific ODI had greater proportion of patients demonstrating response shift than the SF-36. This is likely due to the threshold (MDC) that was required before we considered a real change. The MDC was higher in the generic health outcome questionnaire than in the ODI. On an individual basis, the change in the SF-36-MCS outcome score often did not meet the MDC; hence few patients were documented as having a response shift in the mental component category.

Another factor which may have affected the ability to measure response shift using SF-36-MCS and SF-36-PCS was reported by Busija who showed that in orthopaedic settings the SF-36 showed large to moderate meaningful changes in group scores but low responsiveness to individual change.[121] These authors demonstrated that the SF-36 was best used at a group level to show change in physical, mental, and social dimensions as opposed to being used for
individual patients. In my study, SF-36-PCS for the overall group did have a statistically significant response shift when comparing the pre-test to the then-test. There was no significant response shift at the group level for SF-36-MCS. At an individual level, few patients demonstrated a response shift in the SF-36-MCS outcome measure. This may be explained by the findings of Busija.[121]

Although the current study did not show a response shift for SF-36-MCS, a number of studies in the literature have never-the-less shown a response shift in components of mental health.[19, 58, 62, 153] My inability to demonstrate response shift in the SF-36-MCS may indicate that the SF-36 may not be appropriate due to its poor responsiveness in orthopedic conditions to identify this.

A secondary objective of this study was to determine the level of agreement at two time points for the existence and direction of response shift in individual patients. For the measures of ODI and SF-36-PCS, moderate agreement was demonstrated in the direction of response shift at six weeks and three months post-surgeries. For the SF-36-MCS, the percentage of patients with no response shift, (83 and 78 percent at 6 week and 3 months respectively), provided limited ability to detect response shift (either due to this being less than MDC or the measure lacking responsiveness to change as described by Busija.[121])

Response shift has been documented in a number of health conditions including following total joint replacement surgery.[58, 61] The pattern of response shift that was documented in this study is consistent with other studies of non-terminal and non-malignant disorders. [28, 130] The most common direction in these studies is a retrospective higher estimation of disability prior to treatment. In a total knee replacement population Razmjou found that response shift occurred but there was no ultimate effect on the interpretation of the outcome of patients. In the study by Razmjou, there was no reclassification of patients into a successful outcome after accounting for response shift compared to when not accounting for response shift.[61, 62] Response shift had no clinical significance in that study following total knee arthroplasty. That finding differs from my results in that the outcomes and the interpretation of the treatment effect was influenced by response shift in the present study of a spinal procedure. The lack of any
effect caused by response shift in patients undergoing total joint replacement may be due to the very high efficacy rate of that surgery with a complete treatment effect. Spinal surgery commonly has a less complete treatment effect due to the combination of leg and back pain. The remaining back pain after spinal surgery despite the resolution of the leg pain may contribute to the lower outcome interpretation compared to total joint replacement surgery.[139, 154, 155]

The final objective of this study was to identify potential predictor variables for response shift. The analysis was limited due to the small sample size and no multivariable analysis could be applied. Significant factors were only found for the outcome scale of SF-36-PCS at 3 months. These were age, gender, number of co-morbidities and duration of symptoms. It is difficult to put any clinical significance to this finding as no similar significance was found for the ODI outcome measure. These factors may have predictor relevance and it was my hypothesis that certain predictors could be found that would influence response shift. Further study with a larger sample size is needed to determine if these are predictive of RS in spinal surgery patients.

A number of authors have shown the affect of co-morbidity on a patient’s subjective outcomes, although not taking response shift into account: 1) Slover demonstrated that the number and type of co-morbidity, both medical and psychosocial, impacted the change in both SF-36 and ODI scores after lumbar spine surgery.[125] The average change in bodily pain, physical function, physical component summary scores of the SF-36, as well as ODI scores decreased in response to surgery as the number of co-morbidities increased. Psychosocial co-morbidities such as an active compensation case, self-rated poor health, and smoking also had negative effects on the change in survey scores after surgery. Slover did not suggest any theory to account for the affect on outcome based on psychosocial and medical co-morbidities, but did illustrate the difficulty in isolating true change after spine surgery, and the importance of identifying potential confounders that impact on our ability to measure this. 2) Cella and Tulsky noted that information relating to an individual’s quality of life cannot be abstracted in isolation from coping strategies, past experiences of illness and other dispositional/personality characteristics.[156] 3) Voorhies noted that the results of lumbar decompression are often independent of technical factors.[157] In Voorhies’ study, psychosocial issues including psychiatric diagnosis, compensation claim, and personal injury were predictors of a negative
effect on outcome. In summary, the inter-relationship between patient dispositional or personality factors and outcome are very much intertwined. This is consistent with the model of response shift by Sprangers and Schwartz and was illustrated in Figure 1.[8] The limited positive findings in my study with respect to potential predictors of response shift require further investigation with larger sample size.

An alternate theory to the response shift, known as the "implicit theory of change", has also been suggested as a mechanism to explain how patients make assessments of health over time.[42] Implicit theorists suggest that patients do not remember their initial state and instead extrapolate backwards from their current state. Implicit theory presumes that people begin their recollection by asking themselves how they are currently, followed by asking themselves how they think things have changed, and they then infer what their initial state must have been like. In contrast, response shift theory suggests that the present state differs from a previous state as a result of a new internal standard. This causes a change in the retrospective judgment of the previous state because it is reassessed in a new light based on the new standard. These two theories provide two competing explanations of the same phenomenon.[42]

A further consideration that can affect a patient’s retrospective report of their disability is effort justification bias. This bias would be more implicated in surgical compared to medical interventions for disease. The majority of studies on response shift have been for medical disorders and not for surgical conditions. This bias can lead to a change in the retrospective assessment of preoperative disability in order to self-justify undergoing the ordeal of surgery.[41] This commonly leads to an overestimation of improvement.[36] The impact of effort justification bias is a limitation in this study and may contribute to the direction of response shift seen in the majority of patients, (overestimation of preoperative impairment). Controlling for this bias is difficult. A study with two treatment arms, one surgical and one non-surgical would need to be employed where both treatments are equally effective. Alternatively sham surgery versus real surgery is an alternative way to identify this bias, however an ethically sound study would be difficult to devise.
7.2 Limitations and Strengths

A weakness in this study has been alluded to above; the small sample size. From the inception cohort, 38 patients were lost to follow-up and 19 patients failed to complete the then-test questionnaire. This reduced the sample size to 105 patients. For the analysis of the consistency of the response shift over time, the sample size was further reduced to 63 patients who had outcomes measured at two time points. The excluded patients may have also biased our results, as it is unknown whether these patients were systematically different from those remaining in the study; possibly doing very well and electing not to return or doing very poorly and attending a different hospital. We did test for differences in baseline patient factors between those completing the study and those excluded or not completing the study. There were no significant differences between these groups.

Another limitation is the exclusion criteria used. These limit the external validity of the study. The findings are only applicable to patients with leg dominant pain and not back dominant pain. Spinal degeneration leads to both leg pain and back pain. Leg dominant pain does not imply there is no back pain, however for a given patient, there is a dominance of pain and the current study was for leg dominance. The treatment that was performed, i.e. decompression, is directed at treating only the leg pain. Partial treatment compared to complete treatment is a result and this differentiation may have an influence on response shift.

The use of the then-test for measurement of response shift has been criticized due to the confounding effect of recall bias.[21, 22, 24, 37] The presence or absence of recall bias is a matter of controversy. There are a number of studies that have supported the validity of retrospective data collection for pre-operative disability that indirectly supports the validity of the then-test. Hagg found a high correlation between a patient’s retrospective global assessment of their outcome to a number of prospective outcome measures, (ODI, VAS and Million Score). The global assessment used a retrospective measure of their preoperative disability.[39] In a further study, Howell evaluated patient recall of their preoperative disability following total hip arthroplasty. In this study, QoL status was evaluated at the pre-operative time period by using a questionnaire package that was composed of 3 validated assessment tools: Western Ontario and
McMaster Universities Osteoarthritis Index (WOMAC), the Oxford-12 hip score and the 12-item Short Form score (SF-12). At time periods 3 days, 6 weeks and 12 weeks after surgery the patients completed the same questionnaire package. At 6 weeks post-operatively, there were significant differences observed between the recalled the post-operative scores for WOMAC, Oxford score and SF-12. At the 3 months postoperative time period, there were no significant differences measured by all QoL instruments including WOMAC, Oxford score and SF-12 physical and mental components.[38]

With respect to recall bias, patient recall of their previous level of disability is critical for using the then- test for response shift evaluation. There is controversy about the significance of recall bias in retrospective data collection. In consideration of this controversy, this study could be strengthened if additional methods of measuring RS were used in addition to the then-test looking for convergence of results.

Strength of this study is the use of the MDC. This provided a threshold for each individual patient that eliminated the effect of random error before response shift was considered to be present. No previous study using the then-test for response shift has incorporated the MDC. In our opinion this adds to the validity of using the then-test as a measure of response shift.

### 7.3 Future Directions:

The demonstration of response shift in outcome measurement in longitudinal studies raises questions about the best way to assess outcomes. If the perception of disability changes or the outcome measurement is used differently at two points in time, this difference needs to be captured in the change. For clinicians who treat individual patients, the then-test is a practical method to follow a patient’s trajectory of recovery. Statistical methods that rely on aggregate data such as classification and regression tree analysis or structural equation modeling, are not clinically practical. The Appraisal Method of Rapkin may be a valuable adjunctive method with the then-test. This is a cognitive debriefing using a structured interview with the patient that explores a patient’s perceptions of their QoL and has been used in a study of terminally ill by Westerman.[55] This study found patients’ interpretations of questions changed over the course
of their treatment due to changes in their appraisal processes. The then-test and the Appraisal Method together allow for a quantitative measure of response shift and provide qualitative analysis of what created this shift. Investigating patients’ response shifts and appraisal processes has not been done previously in a population undergoing elective surgery. A study including these measures may also help understand the negative effect on outcome that some specific psychosocial factors (i.e. compensation and litigation) routinely have on outcome.[157]
CHAPTER 3

THESIS DISCUSSION

1.0 Thesis Discussion

This study is the first to demonstrate that response shift is present in a cohort of patients following spinal surgery. In the properly selected patient, outcomes following surgery for spinal stenosis and disc herniation generally show significant improvements at 6 weeks and at 3 months. This study has shown that response shift can influence the effect size of treatment and has implications in the interpretation of outcomes of treatment.

When looking at the overall cohort, the ODI and SF-36-PCS did have a significant response shift. The pattern of response shift in this study was for the majority of patients having a retrospective overestimation of their preoperative disability compared to having an underestimation of their preoperative disability. For the outcome measure ODI, overestimation to underestimation was 52.0% to 20% at 6 weeks and 53.0% to 19.0% at 3 months. For the outcome measure SF-36-PCS overestimation to underestimation was 23.2% to 7.3% at 6 weeks and 40.7% to 13% at 3 months. The proportion of patients with no response shift was too high in the SF-36-MCS outcome measure to make any assessment. The pattern of response shift for ODI and SF-36-PCS is similar to that which has been demonstrated in other non-life threatening conditions. When looking at individual patients, for the ODI and the SF-36-PCS, there was moderate agreement from six weeks to three months in the categorization of response shift (higher retrospective estimation of preoperative disability vs. lower retrospective estimation vs. no response shift). I was unable to comment on the consistency of the response shift for the SF-36-MCS as the number of patients having a response shift was too low based on the threshold of MDC. Furthermore, owing to the few patients who had a response shift in total for SF-36-MCS it is difficult to comment on the existence of response shift using this outcome measure.

The way clinicians evaluate changes in QoL following interventions can have implications regarding the interpretation of the effectiveness of treatment.[15] An adjusted treatment effect
that incorporates response shift will have a larger treatment effect size if a patient has a retrospective greater perception of their preoperative disability. At a programmatic level where decisions must be made about how to expend limited resources, a treatment with a greater benefit is more cost-effective and may attract more resources. Although the adjusted treatment effect has a higher effect size, Howell demonstrated that recalled ratings have higher within-subject error and larger variance, which leads to a reduction in statistical power. These considerations need to be taken into account when designing studies.[38] In my study there was also a high standard error of measurement. This may in part be due to the use of retrospective pre-test data to measure response shift.

The finding in this study of ODI having a greater incidence of response shift compared to the SF-36 may suggest there is a difference in disease specific outcome measures compared to generic outcome measure. This difference however most likely has less to do with the assessment of response shift but rather to the sensitivity to change for the outcome instrument. The MDC for the SF36 subscales was high (8 and 14 for SF-36-PCS and SF-36-MCS respectively). This created a high threshold to obtain before it was considered statistically significant for a response shift as opposed to the change being within statistical error. Busija showed that in orthopaedic settings the SF-36 showed large to moderate meaningful changes in group scores but low responsiveness to individual change.[121]

Response shift can occur from any change in health status. It has been discussed above how this has implications on our assessment of outcome. Response shift is a passive effect of treatment and on our measurement abilities of the outcomes of treatment. However utilizing a patient’s response shift can also be used as an active method to facilitate rehabilitation and/or adapt to a chronic disease. This has been employed in a population of patients suffering from a stroke.[30, 58] Teaching patients cognitive strategies for adapting to their change in health status can occur by changing values and conceptualizations. For patients undergoing surgery, surgeons can ensure that patients have reasonable expectations about the outcomes of surgery which may help facilitate a favorable response shift and increase the effectiveness of surgery.
Whether we regard the “true” treatment effect as the standard pre-post difference, or the adjusted difference (then test-post difference), ultimately it is the patient’s perception of the change. It is the hypothesis of this thesis that the true perception of change is the adjusted treatment effect as a response shift can occur causing a change in the way a patient evaluates his/her disability or QoL over time. This study has demonstrated that response shift occurs and will have implications on the interpretation of treatment effectiveness. Other measures of response shift may be useful on a go forward basis. The Appraisal Method of Rapkin may be valuable to further define the patient factors that mediate response shift and may be beneficial in helping to predict prior to an intervention that will be susceptible to a response shift and the direction it is likely to take.[7]

The patient’s perception of change is a matter of importance in the ongoing development and psychometric validation of outcome questionnaires. In spinal surgery the influence of response shift should be considered by those stakeholders who have an interest in the evaluation of outcome from a surgical intervention.
References


89. Parent, E., Moffet, H., Comparative responsiveness of locomotor tests and questionnaires used to follow early recovery after total knee arthroplasty, Arch Phys Med Rehabil. 2002 Jan;83(1): 70-80.


116. Ware JE, S.K., Kosinski M, et al., SF-36 Health Survey Manual and Interpretation Guide. 1993, Boston MA: New England Medical Center, the Health Institute,


APPENDIX 1: OSWESTRY FUNCTIONAL ASSESSMENT QUESTIONNAIRE

Please Read:
This questionnaire has been designed to give us information as to how your pain has affected your ability to manage in everyday life. Please answer every section and mark in each section only the one box that applies to you. We realize you may consider that two of the statements in any one section relate to you, but please just mark the box that closely describes your problem.

Section 1: Pain Intensity
- I can tolerate the pain I have without having to use painkillers.
- The pain is bad, but I manage without painkillers.
- Painkillers give complete relief from pain.
- Painkillers give moderate relief from pain.
- Painkillers give very little relief from pain.
- Painkillers have no effect on the pain and I do not use them.

Section 2: Personal Care (washing, dressing, etc.,)
- I can look after myself normally without causing extra pain.
- I can look after myself normally but it causes extra pain.
- It is painful to look after myself and I am slow and careful
- I need some help but manage most of my personal care.
- I need help every day in most aspects of my personal care.
- I do not get dressed, wash with difficulty and stay in bed.

Section 3: Lifting
- I can lift heavy weights without extra pain.
- I can lift heavy weights but it gives extra pain.
- Pain prevents me from lifting heavy weights off the floor, but I can manage if they are conveniently positioned.
- Pain prevents me from lifting heavy weights but I can manage light to medium weights if they are positioned conveniently.
- I can only lift very light weights.
- I cannot lift or carry anything at all.

Section 4: Walking
- Pain does not prevent me walking any distance.
- Pain prevents me walking more than 1 mile.
- Pain prevents me walking more than ½ mile.
- Pain prevents me walking more than ¼ mile.
- I can only walk using a stick or crutches.
- I am in bed most of the time and have to crawl to the toilette.

Section 5: Sitting
- I can sit in any chair as long as I like.
- I can only sit in my favorite chair as long as I like.
- Pain prevents me sitting more than 1 hour.
- Pain prevents me sitting more than ½ hour.
- Pain prevents me sitting more than 10 mins.
- Pain prevents me sitting at all.

Section 6: Standing
- I can stand as long as I want without extra pain.
- I can stand as long as I want but it gives me extra pain.
- Pain prevents me from standing for more than 1 hour.
- Pain prevents me from standing for more than 30 mins.
- Pain prevents me from standing for more than 10 mins.
- Pain prevents me from standing at all.
Section 7: Sleeping
- Pain does not prevent me from sleeping well.
- I can sleep well only by using tablets.
- Even when I take tablets I have less than six hours sleep.
- Even when I take tablets I have less than four hours sleep.
- Even when I take tablets I have less than two hours sleep.
- Pain prevents me from sleeping at all.

Section 8: Sex Life
- My sex life is normal and causes no extra pain.
- My sex life is normal but it increases the degree of pain.
- My sex life is normal but is very painful.
- My sex life is severely restricted by pain.
- My sex life is nearly absent because of pain.
- Pain prevents any sex life at all.

Section 9: Social Life
- My social life is normal and gives me no extra pain.
- My social life is normal but increases the degree of pain.
- Pain has no significant effect on social life apart from limiting my more energetic interests, e.g. Dancing, etc.,
- Pain has restricted my social life and I do not go out as often.
- Pain has restricted my social life to my home.
- Pain prevents any social life at all.

Section 10: Travelling
- I can travel anywhere without extra pain.
- I can travel anywhere but it gives me extra pain.
- Pain is bad but I can manage journeys over two hours.
- Pain restricts me to journeys of less than one hour.
- Pain restricts me to short necessary journeys under 30 mins.
- Pain prevents me from travelling except to the doctor or hospital
APPENDIX 2 - THE MOS 36-ITEM SHORT-FORM HEALTH SURVEY (SF-36)

INSTRUCTIONS: This survey asks you for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:
   (circle one)
   1. Excellent . . . . . . . . . . . . . . . . . . .
   2. Very Good . . . . . . . . . . . . . . . . . . .
   3. Good . . . . . . . . . . . . . . . . . . . . .
   4. Fair . . . . . . . . . . . . . . . . . . . . .
   5. Poor . . . . . . . . . . . . . . . . . . . . .

2. Compared to one year ago, how would you rate your health in general now?
   (circle one)
   1. Much better now than one year ago . . . . . .
   2. Somewhat better now than one year ago . . . . .
   3. About the same as one year ago. . . . . . . . .
   4. Somewhat worse now than one year ago . . . . .
   5. Much worse now than one year ago. . . . . . .
3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

(Circle one number on each line)

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Yes, Limited A Lot</th>
<th>Yes, Limited A Little</th>
<th>No, Not Limited At All</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c. Lifting or carrying groceries</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>d. Climbing several flights of stairs</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>e. Climbing one flight of stairs</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>f. Bending, kneeling, or stooping</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>g. Walking more than a mile</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>h. Walking several blocks</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>i. Walking one block</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>j. Bathing or dressing yourself</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(Circle one number on each line)

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the <strong>amount of time</strong> you spent on work or other activities</td>
<td>1</td>
</tr>
<tr>
<td>b. <strong>Accomplish less</strong> than you would like</td>
<td>1</td>
</tr>
<tr>
<td>c. Were limited in the <strong>kind</strong> of work or other activities</td>
<td>1</td>
</tr>
<tr>
<td>d. Had <strong>difficulty</strong> performing the work or other activities (for example, it took extra effort)</td>
<td>1</td>
</tr>
</tbody>
</table>
5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the <strong>amount of time</strong> you spent on work or other activities</td>
<td>1</td>
</tr>
<tr>
<td>b. <strong>Accomplish less</strong> than you would like</td>
<td>1</td>
</tr>
<tr>
<td>c. Didn’t do work or other activities as <strong>carefully</strong> as usual</td>
<td>1</td>
</tr>
</tbody>
</table>

6. During the past week, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?

(circle one)

- Not at all
- Slightly
- Moderately
- Quite a bit.
- Extremely

7. How much **bodily** pain have you had during the past week?

(circle one)

- None
- Very mild.
- Mild
- Moderate
8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

- Not at all
- A little bit
- Moderately
- Quite a bit
- Extremely

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past week –

(circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>All of the time</th>
<th>Most of the time</th>
<th>A Good Bit of the time</th>
<th>Some of the time</th>
<th>A Little of the Time</th>
<th>None of the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Did you feel full of pep?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Have you been a very nervous person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Have you felt so down in the dumps that nothing could cheer you up?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Have you felt calm and peaceful?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>e. Did you have a lot of energy?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
f. Have you felt downhearted and blue?  

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
</table>


g. Did you feel worn out?  

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
</table>


h. Have you been a happy person?  

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
</table>


i. Did you feel tired?  

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
</table>


10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?  

(circle one)  

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All of the time</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Most of the time</strong></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some of the time</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A little of the time</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None of the time</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


11. How TRUE or FALSE is each of the following statements for you?  

(circle one number on each line)  

<table>
<thead>
<tr>
<th></th>
<th>Definitely True</th>
<th>Mostly True</th>
<th>Don’t Know</th>
<th>Mostly False</th>
<th>Definitely False</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I seem to get sick a little easier than other people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. I am as healthy as anybody I know</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. I expect my health to get worse</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. My health is excellent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
APPENDIX 3 – RESEARCH ETHICS BOARD APPROVAL

RENEWED ETHICS APPROVAL

RESEARCH ETHICS BOARD
SUNNYBROOK
HEALTH SCIENCES CENTRE
S133, 2075 Bayview Avenue
Toronto, Ontario
M4N 3M5

Must be completed annually for all on-going studies
Please return to room S133

Project Title: Response Shift in Patient Expectations and Functional Outcome Following Posterior Lumbar Spinal Surgery for Degenerative Spinal Disorders

Project Identification Number: 151-2004

Original Approval Date: June 10, 2004

Principal Investigator: Dr. Joel Finkelstein

Full Address Including Room Number: Sunnybrook Health Sciences Centre
2075 Bayview Av., RM MG-361
Toronto ON M4N 3M5

Full Board Review Required: Yes ☐ No ☒

If industry sponsored, please note there is a $500 re-approval fee. Invoicing information including contact name and full mailing address is required for each renewal.
Sunnybrook Health Sciences Centre
Research Ethics Board
Renewed Ethics Approval

Progress of Study (provide a brief summary of patient accrual):

Since the last renewal, 145 new patients have been consented. These patients have been followed for the past year with 31 of them being lost to follow up. The remaining 114 patients have successfully completed the Short Form 36 (SF-36) and the Oswestry Disability Questionnaire at the appropriate time points and continue to be followed. There have been no protocol deviations.

Amendments to the study (must be submitted for approval):

None

Changes in scientific knowledge that could impact on the study and action taken:

N/A

Unexpected or adverse events and action taken:

N/A

Protocol violations and actions taken:

None

Expected date of completion:

May 2009
Sunnybrook Health Sciences Centre
Research Ethics Board
Renewed Ethics Approval

Please sign below:

My signature certifies the following information is correct and I will not use any procedures, which have not been approved by the Board.

Signature of Principal Investigator

Signature

Date

April 30, 2008

The above study is ethically acceptable and has received renewed ethics approval. This study may continue at Sunnybrook Health Sciences Centre.

Philip C. Hébert MD, PhD. FCFP
Chair, Research Ethics Board

Date of Review

May 2, 2008

Date of Full Board Review (if required):
APPENDIX 4 – LIFE SATISFACTION QUESTIONNAIRE

☐ Pre-op
☐ Post-op
☐ Then-test (Renewed judgment)

Here are a number of statements concerning how satisfied you are with different aspects of your life. For each of these statements please mark a number from 1 to 6, where 1 means very dissatisfying and 6 very satisfying.

1: Very Dissatisfying
2: Dissatisfying
3: Rather Dissatisfying
4: Rather Satisfying
5: Satisfying
6: Very Satisfying

1. My life as a whole is: 1 2 3 4 5 6
2. My vocational situation is: 1 2 3 4 5 6
3. My financial situation is: 1 2 3 4 5 6
4. My leisure situation is: 1 2 3 4 5 6
5. My contact with friends and acquaintances is: 1 2 3 4 5 6
6. My sexual life is: 1 2 3 4 5 6
7. My ability to manage my self-care (dressing, hygiene, transfers,...) is: 1 2 3 4 5 6
8. My family life is:
☐ N/A, (I have no family) 1 2 3 4 5 6
9. My partner relationship is
☐ N/A, I have no steady partner relationship 1 2 3 4 5 6
10. My physical health is: 1 2 3 4 5 6
11. My psychological health is: 1 2 3 4 5 6