Predictors of Short-Term Patient-Reported Outcome Following Surgical Treatment of Rotator Cuff Pathology

by

Jason D. Woollard

Bachelor of Arts, Ohio Wesleyan University, 1997

Master of Physical Therapy, Hahnemann University, 1999

Submitted to the Graduate Faculty of the

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

School of Health and Rehabilitation Sciences

This dissertation is presented

by

Jason D.Woollard

It was defended on

April 17th, 2012

and approved by:

James E. Bost, MS, PhD, Associate Professor, School of Medicine

G. Kelley Fitzgerald, PhD, PT, FAPTA, Associate Professor, SHRS

Sara Piva, PhD, PT, Assistant Professor, SHRS

Mark Rodosky, MD, Assistant Professor, School of Medicine

Chair of Dissertation Committee: James J. Irrgang, PhD, PT, ATC, FAPTA, Associate Professor, Department of Orthopaedic Surgery, School of Medicine

Predictors of Short-Term Patient-Reported Outcome Following Surgical Treatment of Rotator Cuff Pathology

Jason D. Woollard PhD, MPT

University of Pittsburgh, 2012

The purpose of this dissertation was to systematically review the prognostic evidence for factors that may predict clinical outcome in individuals undergoing rotator cuff repair, determine preoperative factors that can accurately predict outcome in individuals having arthroscopic subacromial decompression with or without rotator cuff repair, and calculate responsiveness for the Western Ontario Rotator Cuff Index and the Disabilities of the Arm, Shoulder and Hand Questionnaire in these individuals.

A preoperative evaluation collected demographic information, history of the shoulder condition, measures of shoulder impairment, shoulder activity level, fear-avoidance levels, depressive symptomatology, and anxiety. Patient-reported outcomes (PROs) included a disease-specific PRO, the Western Ontario Rotator Cuff Index (WORC), and a region-specific PRO, the Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH). Six months postoperatively, the WORC, DASH and global rating of change were collected.

Logistic regression analysis determined which preoperative variables were able to predict responders from nonresponders. Responders were determined based on a minimum improvement of 17-points on the WORC score and a global rating of change score of at least "quite a bit better" at the 6-month postoperative time point. Linear regression, with the WORC change score used as the dependent variable, provided a secondary analysis to allow comparison of the logistic and linear models. Effect sizes, standardized response means and the sensitivity

and specificity of the minimal clinically important difference for both the WORC and DASH were calculated.

Surgery on the dominant shoulder and a score of 25 or less on the work subscale of the Fear-Avoidance Beliefs Questionnaire were the significant predictors in the final logistic model. The accuracy of the model for correctly predicting responders from nonresponders was excellent. Fear-avoidance, as a predictor of outcome, provides a modifiable factor that can be targeted by specific rehabilitation interventions. In the linear model, the WORC change score was predicted by surgery on the dominant arm, modified job duty, and age. Both the WORC and DASH demonstrated high levels of internal responsiveness while external responsiveness could not be accurately assessed due to the preponderance of responders to nonresponders.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

During the course of my doctoral training, I have been fortunate to work with so many excellent people. From Dr. James Irrgang as my primary mentor, to my fellow doctoral students, I am lucky to have crossed the paths of so many motivated, intelligent and dedicated researchers, educators, clinicians, and students. The amount of knowledge and insight that I have gained from all of you is incredible.

My committee's commitment to my training has been exemplary. Dr. James Irrgang taught me almost everything I know about evidence-based practice. His mentorship allowed me to accept a faculty position and be ready to teach evidence-based practice to aspiring physical therapists. For my dissertation, Jay has spent many, many hours reading my work and providing excellent feedback which has allowed me to complete this project successfully and develop as both a researcher and educator. He has always made time for me and I am grateful for his support.

Dr. Kelley Fitzgerald and Dr. Sara Piva allowed me to work with them on several research projects that helped me develop the skills necessary to carry out my own research. Their inclusion of me into their research projects as well as their thoughtful and constructive feedback throughout the years has allowed me to publish manuscripts that would not have come to fruition without their support.

Dr. James Bost, with his extensive knowledge of statistics, always made time for me to meet with him and even offered his expertise to my project after leaving the University of Pittsburgh. His commitment to my work is greatly appreciated.

Dr. Mark Rodosky, whose excellent surgical outcomes ruined my manuscript on minimally detectable change for the WORC and DASH, allowed me to clog up his examination rooms for the sake of this research. His willingness to allow me to recruit his patients for this project was instrumental in the completion of this research.

Being able to work with and forge friendships with Lucas Simoes, Dr. Ale Gil, Paulo Teixeira, Gustavo Almeida, Ibrahim Altubasi, and many others was truly a blessing. I am fortunate to have you as my friends.

To my parents, Larry and Teresa Woollard, whose personal sacrifices allowed me to be in a position to complete a doctoral degree; I could not have asked for better parents.

Finally, the main reason I was able to complete this degree was because of my incredible wife, Dr. Fabrisia Ambrosio Woollard. She supported me, encouraged me, and pushed me when needed in order to finish this degree. Her chameleonic ability to be a pioneering researcher, an incredible mother and a great wife has inspired me. I know that without her, this accomplishment would not have happened. I am so fortunate to have her in my life.

1.0 INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

When establishing the prognosis for surgical intervention of a patient's rotator cuff lesion evidence-based prognostic evidence should be used to formulate the prognosis. While many outcomes studies, case-series, and a handful of randomized trials have shown relatively positive outcomes for individuals undergoing rotator cuff repair of small cuff tears or subacromial decompression for subacromial impingement syndrome, there are individuals who undergo these procedures that fail to respond favorably.[1-18] By understanding the preoperative factors that can be used to estimate an individual's response to rotator cuff repair and subacromial decompression, individuals involved in the decision to have elective shoulder surgery can combine the best empiric evidence and surgeon experience to make a clinical-decision. If preoperative factors are identified that increase or decrease the likelihood of successful response to surgery, then an evidence-based prognosis of outcome can be provided and discussed with the surgical candidate. For instance, if the surgical candidate has factors that improve the probability of a successful outcome, the prognostic evidence could solidify the decision to have surgery. If the surgical candidate has preoperative factors that have been shown to decrease the probability of a successful outcome, then this information can be used to temper expectations if it is determined that surgery is still the best option. Alternatively, if the patient has factors that predict a poor outcome from surgery perhaps they may benefit from a treatment approach that does not include surgery**.** In addition to aiding the clinical decision making process, modifiable predictors of outcome should be considered as targets for future intervention trials. For instance, future controlled trials could examine the impact of either preoperative or postoperative rehabilitation treatments that specifically target the robust predictors of outcome in this patient population. Since physical therapists routinely treat this patient population prior to and following surgical intervention, it is important that modifiable factors are considered for their effect on outcome and that research be carried out to determine if interventions targeting the modifiable factors result in improved outcomes.

While there is literature that has examined the predictors of outcome in these patients, many of the studies demonstrate limitations such as outcome measures that lack validation, statistical analyses that do not include multiple, possibly competing predictors, and/or retrospective data collection. Therefore, additional research that uses sound methodology and determines outcome using a valid and responsive patient-reported outcome (PRO) would be beneficial for strengthening the evidence of preoperative factors that can be used to distinguish responders from nonresponders in this population. In addition, many factors that potentially influence response in rotator cuff surgery have yet to be investigated for their prognostic abilities. These include psychosocial factors, scapular dyskinesis, glenohumeral internal rotation deficits, and shoulder strength measured using clinically available handheld dynamometry. All of these factors are clinically modifiable and could be addressed during physical therapy.

Since the patient's perception of outcome is arguably the most important metric of successful shoulder surgery, PROs used to establish the criteria for determining responders from nonresponders should include valid instruments for measuring health. The World Health

Organization defines health as, “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.”[19] Therefore, PROs used in the determination of who is a responder to rotator cuff surgery and who is not should incorporate multi-dimensional aspects of health. Both the Western Ontario Rotator Cuff Index and the Disabilities of the Arm Shoulder and Hand Questionnaire are PROs that are valid, reliable and incorporate the multi-dimensional aspects of health.[20-27]

Therefore, the aims of this dissertation are to examine the current state of the evidence for predicting clinical outcome in individuals requiring rotator cuff repair, to evaluate preoperative predictors of response to surgery and determine which predictors can accurately discriminate responders from nonresponders after rotator cuff surgery, and to directly compare two PROs, the Western Ontario Rotator Cuff Index and the Disabilities of the Arm Shoulder and Hand Questionnaire, for their responsiveness, and to determine the minimal clinically important difference for these 2 PROs.

1.2 AIMS AND HYPOTHESES

1.2.1 Specific Aim 1

To systematically review the current body of literature that examines the relationships between preoperative patient data and patient reported outcomes in individuals requiring surgical intervention for a rotator cuff tear

1.2.1.1 Hypothesis 1

Significant limitations exist in published studies that examine preoperative predictors of outcome in individuals undergoing rotator cuff repair. These limitations will be due to choices in study design, statistical analyses of the data, and the instrument chosen to measure outcome of the surgery

1.2.2 Specific Aim 2

To examine preoperative predictors collected during an examination for their prognostic accuracy in predicting responders of surgery from nonresponders in individuals requiring arthroscopic subacromial decompression with or without rotator cuff repair

1.2.2.1 Hypothesis 2

A set of preoperative predictors will accurately distinguish responders from nonresponders after arthroscopic subacromial decompression with or without rotator cuff repair

1.2.3 Specific Aim 3

To calculate the minimum clinically important difference for the Western Ontario Rotator Cuff Index (WORC) and the Disabilities of the Arm Shoulder and Hand Questionnaire (DASH) and to compare the responsiveness of these patient-reported outcomes

1.2.3.1 Hypothesis 3

The WORC, a disease-specific measure, will demonstrate greater internal and external responsiveness than the DASH, a region-specific measure

2.0 PREDICTORS OF CLINICAL OUTCOME IN ROTATOR CUFF REPAIR: A SYSTEMATIC REVIEW

2.1 INTRODUCTION

In the 2008 report published by the American Academy of Orthopedic Surgeons, *The Burden of Musculoskeletal Disease in the United States*, shoulder pain was the 2nd most prevalent self-reported musculoskeletal condition.[28] The report also indicates that lesions of the rotator cuff can be expected to increase in the future given the aging of the American population. [28] Over the first decade of the 21st century, the frequency of rotator cuff repair grew more rapidly in comparison to other orthopedic surgical procedures. In an analysis of the case mix of candidates preparing for certification by the American Board of Orthopedic Surgery, Garrett et al[29] demonstrated that the volume of rotator cuff repair repairs grew rapidly from 1999 to 2003. In comparison to the remaining 24 orthopedic procedures that comprised the 25 most common procedures performed by the candidates for board certification, CPT code 23412, repair of tendon(s)/rotator cuff, experienced the 2nd largest increase in rank among top CPT codes from 37th in 1999 to 14th in 2003. CPT 27245, repair of thigh fracture/trochanteric, was the only CPT code to experience a larger jump in rank of procedures performed over that timeframe. A 2012 study examined the American Board of Orthopedic Surgery case mix for fellowship candidates and found a practice pattern shift towards arthroscopic repairs of the rotator cuff with and without subacromial decompression as well as a shift away from isolated subacromial decompression.[30]

With the expected increase in volume of rotator cuff repairs, the ability to accurately predict, preoperatively, the probability of successful rotator cuff surgery would provide physicians and their patients with valuable information when making shared decisions regarding surgical intervention. Accordingly, increased knowledge of the impact of preoperative factors would improve the patient's understanding of what factors increase or decrease the probability of achieving a successful surgical outcome.

Previous studies of preoperative predictors of the outcome of rotator cuff repair have focused on the relationship of age[2, 6, 10, 12, 14, 16, 31-33], sex[4, 6, 9, 14, 32, 34], worker's compensation status [4, 5, 35-38], and fatty infiltration of the rotator cuff muscles [4, 7, 12, 14, 32, 38]. Due to the lack of consensus on the most appropriate clinical outcome instruments to use in studies regarding rotator cuff repair, these studies used a wide variety of clinical outcomes. Additionally, other studies that have examined preoperative predictors such as shoulder motion[16], shoulder strength[13], comorbidities [3, 37, 39], involvement of the dominant arm [4, 14, 15, 38], duration of symptoms [4, 10, 14, 16, 31, 38], shoulder activity level [4, 15], and smoking history. [15, 34, 38] While research examining the predictors of rotator cuff repair outcome has been done, it is difficult to assess the quality and emerging themes from these studies without examining them systematically. By systematically examining these studies, the strengths and weaknesses in the current evidence may become more apparent. A better understanding of the preoperative predictors of clinical outcome after rotator cuff repair may lead to predictive models that can be used to inform patients about their probability of a successful repair given their unique characteristics. A tool such as this will allow patients to make an informed evidence-based decision on whether or not to have elective surgery for their rotator cuff pathology. The purposes of this systematic review are to: 1.examine and appraise the current evidence related to preoperative predictors of clinical outcome of rotator cuff repair and 2. recommend future steps that can be taken to improve the utility of prognostic evidence for the outcome of rotator cuff repair.

2.2 METHODS

2.2.1 Search Strategy

To identify relevant literature, Medline was searched July 2011 and again November 2011. The following keywords [affect OR determinants OR effects OR factors OR impact OR indicators OR influence OR predict] were combined with the MeSH term, “Prognosis” using the Boolean connector “OR” to find articles focused on prognosis and prediction. This combination was then combined by the Boolean operator “AND” to the following search strategy, [Rotator cuff/surgery (MeSH)] to identify articles related to prognosis or prediction of outcome of rotator cuff surgery. This search strategy retrieved 938 articles (Figure 1). Embase was also searched in November 2011 using the keywords ‘rotator cuff surgery’ and ‘prognosis’ and combined by the Boolean connector “AND”. In addition to the database searches, the bibliographies of articles that were retained and abstracted were searched for additional pertinent articles.

2.2.2 Article Selection Criteria

The articles retrieved through Medline and Embase were added to a matrix spreadsheet if they appeared to meet the title and abstract requirements listed in Table 1. In order to examine the methodology of each study, the Matrix Method[40], was used to record the study design (prospective/retrospective; single/multi-site; derivation/validation study), the number of subjects, the specific clinical outcome instrument used as the dependent variable (impairment-based/patient reported), the statistical methods used to analyze the relationship between the potential predictor variable and the postoperative clinical outcome score, control of confounders, assessment of model strength, and the attrition rate. Upon review of the retrieved articles, it was decided to use articles published since 1995 in an attempt to focus the review on more current surgical techniques such as arthroscopic rotator cuff repair and attempt to minimize the number of studies that utilized open cuff repairs.

2.2.3 Study Quality

The evaluation of study quality was based on the recommendations of Beattie and Nelson[41], Hayden et al[42], and Hess [43]. Table 2 lists the 7 criteria used to quantify study quality as well as the score each study received for the 7 criteria. Each criterion was scored as either "yes" or "no". If a study fulfilled the criterion, it was scored a "yes". If the study did not fulfill the criterion, it was scored a "no". Each "yes" response counted for 1-point and the number of "yes" responses were summed to provide a final score of study quality that ranged from 0 to 7 (with higher scores indicating better study quality). It is important to note that a gold standard for assessing the quality of prognostic studies does not exist. Therefore, the quality score given in this review is based on recommendations of the authors noted above and has limitations.

The criteria used to quantify study quality included the study's design, attrition rate, validation in a second sample, and statistical analyses. For study design, studies that were prospective rather than retrospective were awarded 1-point, based on the fact that retrospective studies may be susceptible to greater recall bias and the fact that prospective studies allow greater control of data collection[43, 44] A multi-site study, as opposed to a single site of data collection, was awarded a quality point given the increased external validity provided by a multi-site study design. Prospective studies added an additional quality point if their attrition rate was less than 20% of the baseline sample. Attrition rates can affect prognostic accuracy and high attrition rates potentially lead to inflated estimates of positive outcomes[42, 45] Due to the importance of validating the findings of a prognostic study in a second, independent sample a study was given a quality point if it was a validation study as opposed to a derivation study.[46, 47]**.** Finally**,** 3 of the possible 7 quality points were based on criteria relating to the study's statistical analyses and presentation of the statistical analyses. If the strength of the relationship between the predictor and the outcome was reported, a point was awarded. If the strength of the predictor was reported after controlling for potential confounding variables, another point was awarded. If the strength of the predictive model was reported (R^2, Pseudo R^2, Area under the ROC curve) another point was awarded. [41]

2.3 RESULTS

From the Medline search via PubMed, 938 studies were retrieved and the titles and abstracts of these articles were examined to determine if they potentially meet the inclusion criteria for this systematic review. Upon review, 40 of the 938 studies whose titles and/or abstracts indicated the study examined predictors of the clinical outcome of rotator cuff repair were added to the matrix for complete review of the article. In addition to these 40 articles, 16 more articles were identified by review of the reference lists of the original 40 articles**.** Of the 56 predictor articles reviewed in the matrix, 35 were subsequently eliminated for not meeting the criteria for final inclusion in Table 1. This resulted in 21 articles being utilized for this systematic review.
Figure 1 is a flowchart summarizing the process of article identification and selection. From the EMBASE search, 160 articles were retrieved, but all of the articles that appeared to meet the inclusion/exclusion criteria (n=18) were either duplicates of the Medline search, written in a foreign language, or did not meet all the criteria for inclusion in the review. Of the 21 studies meeting the criteria for inclusion in this systematic review, 8 predictors were investigated in at least 3 studies and included age, sex, fatty infiltration of the rotator cuff musculature, workman's compensation status, smoking status, comorbidities, involvement of dominant arm, and duration of symptoms.

2.3.1 Age

Eleven studies examined the relationship of age to clinical outcome following rotator cuff repair. (Table 3). Eight of these studies were prospective studies and 3 were retrospective studies. Of the 8 prospective studies, 5 found a correlation of age to clinical outcome while 3 did not. Two of these studies used multiple linear regression (MLR) and 1 used logistic regression to determine if age was an independent predictor of outcome given other predictors were in the model. Oh et al[15], although reporting a statistically significant positive correlation of age to the Constant score, found upon MLR analyses that age was not an independent predictor of the Constant score after controlling for abduction torque of the uninvolved shoulder. Gulotta et al also found age was not a predictor of an American Shoulder and Elbow Surgeons score of greater than 90.[8] Milano et al[11] also utilized MLR to examine predictors of outcome and, contrary to the results of Oh et al[14], age remained an independent predictor of the Disability of the Shoulder, Arm, Hand (DASH) score after controlling for shape of the tear, retraction and reducibility of the tear, fatty degeneration of the cuff musculature, involvement of the subscapularis, and repair technique.

From the 3 retrospective studies, Bjornsson et al[31] used t-tests to compare the Constant score, DASH and Western Ontario Rotator Cuff Index scores of those subjects older than 65 years of age to those less than 65 years old. They found no difference for these outcome measures between the age groups. Hersch et al[10] reported older age was significantly associated with a better outcome on the American Shoulder and Elbow Surgeons (ASES) score and UCLA scores, but age demonstrated no association with the Constant score. When they used MLR to analyze the effect of age on the ASES and UCLA scores, age failed to remain an independent predictor of either score. In contrast to the results of Hersch et al[10], Oh et al[14] did find a correlation between increasing age and worsening Constant scores. However, when MLR was used to examine the potential predictors of the outcome, age was no longer an independent predictor of any of the outcomes. Overall, the results are inconsistent with regards to the predictive ability of age. While a majority of the studies reported a correlation of age to outcome, 4 of the 5 studies that used regression to control for possible confounding variables reported that age was not predictive of outcome. However, Milano et al received the highest quality score and did report that increasing age resulted in worse DASH scores following rotator cuff repair. While inconsistent as a predictor, age does not predict better outcomes and may predict worse outcomes on the DASH.

2.3.2 Sex

Three prospective cohort studies and 4 retrospective studies examined the relationship of sex to clinical outcomes (Table 4). In the prospective studies, Gartsman et al[6] used Wilcoxon rank sum tests and found no difference between sexes for any of three clinical outcomes. Charousset et al[4] used MLR and found sex to be an independent predictor of worse Constant scores for females. In the prospective cohort study by Oh et al[14], being female correlated with worse SST scores but not with the Constant or ASES scores. In their methods, the authors state they carried out a multiple regression analysis, but they did not present regression findings.

In the 4 retrospective studies, 2 studies[34, 38] found no statistically significant differences in clinical outcome between men and women using t-tests to compare mean differences. One study used correlations of sex to outcome score and also found no statistically significant relationship[15]. The 4th retrospective study by Henn et al[9] found patient sex was not a predictor of SST scores or DASH scores using MLR. Based on these studies, a preponderance of the evidence indicates gender is not a consistent predictor of clinical outcomes following rotator cuff repair.

2.3.3 Fatty Infiltration

Goutallier et al established criteria for grading the amount of fatty infiltration present in the rotator cuff muscles.[7] Their grading system, the global fatty degeneration index (GFDI), for the amount of fatty infiltration within a rotator cuff muscle ranges from zero indicating no fatty deposits to 4 indicating more fatty infiltrates are present than muscle tissue. Charousset et al[4] ,Oh et al[14], and Milano et al[11] all conducted prospective studies and used MLR to determine whether or not fatty infiltration of the rotator cuff musculature predicted worse clinical outcome scores. In the first two studies, the GFDI was not a predictor of clinical outcome. In the study by Milano et al[11] , the GFDI was an independent predictor of the postoperative work-DASH score but not the overall DASH score.

Three retrospective studies also evaluated the effects of rotator cuff fatty infiltration on clinical outcome. In their 2010 study, Oh et al[15] again analyzed the predictive ability of fatty degeneration and again found it not to be a predictor of any of the clinical outcome measures that were investigated. Petersen and Murphy[38], in a small (n=36) retrospective study, also found GFDI to not be predictive of clinical outcome, as determined by the ASES or the UCLA. The remaining study by Goutallier et al[7] did find that the GFDI correlated with the subjects' Constant scores. From the studies reviewed, a preponderance of the evidence found the GFDI does not predict clinical outcomes.

2.3.4 Worker's Compensation

Charousset et al[4] and Henn et al[35] are both prospective cohort studies that used MLR to determine the predictive value of a worker's compensation (WC) claim. These studies used

MLR to analyze the strength of the predictors in the presence of potential confounders, making these two studies the strongest of the 6 studies examined. However, the study by Charousset et al[4] only had 12 of 114 subjects who were WC claimants and the outcome measure used was the Constant score[4]. They reported no relationship between the presence of a work-related injury and outcome of the rotator cuff repair. The study by Henn et al[35]had 39 workers compensation claimants and utilized the DASH as the dependent variable for the MLR analysis. In this study, having a worker's compensation claim was an independent predictor of a worse DASH score.

The studies by McKee et al[36] and Cole et al[5] were both prospective cohort studies that compared average outcome scores to determine if there was a difference in the clinical outcome for individuals with a WC claim as compared to those without a claim. The McKee et al study[36], using the Shoulder Pain and Disability Index (SPADI) as the outcome measure, reported statistically significant differences between the WC and non-WC groups. On the other hand, Cole et al[5] found no differences between the WC and non-WC groups for the Constant, ASES, SST or Rowe Score.

Namdari et al[37] and Petersen and Murphy[38] both conducted retrospective studies. The study by Namdari et al included 43 subjects with WC claims and utilized the DASH and change in DASH as the dependent variables in MLR analyses. They found that WC status was an independent predictor of the DASH but not the change in DASH after controlling for total comorbidities, obesity, and diabetes mellitis. The study by Petersen and Murphy included only 12 subjects with a WC claim and reported no statistically significant difference (using t-tests) for the ASES and UCLA scores of WC claimants when compared to non-WC claimants.

Having a worker's compensation claim was never a predictor of improved outcome and the studies of higher quality indicated that having a worker's compensation claim predicted less successful clinical outcome when the outcome was the DASH. However, having a worker's compensation claim was not a predictor of the DASH change score in the study by Namdari et al even when it did predict the DASH score in the same sample. Worker's compensation did not predict Constant scores. Therefore, it predicted outcome when outcome was measured by the DASH, but not when change in DASH score or the Constant score were used to measure outcome.

2.3.5 Smoking

Three studies examined the relationship of smoking to clinical outcomes following rotator cuff repair. All were retrospective in design. Mallon et al[34] reported smoking status to be a significant predictor of the UCLA score. Oh et al[15] examined the relationship of smoking to the Constant score and found no relationship between the 2 variables. Petersen and Murphy[38] also found no relationship between smoking and ASES or UCLA scores. The low quality scores for the 3 studies reviewed indicates additional studies are appropriate to better elucidate the predictive ability of smoking status.

2.3.6 Comorbidities

Three studies examined the relationship of medical comorbidities to clinical outcomes following rotator cuff repair. Two prospective cohort studies, Boissonault et al[3] and Tashjian et al[39] reported no relationship between the number of comorbidities and clinical outcomes following rotator cuff repair. Both studies used MLR to analyze this relationship. Namdari et al[37], in a retrospective case-control design study, reported a correlation of 0.36 ($p<0.01$) between the total number of comorbidities and the postoperative DASH score. The studies with stronger quality indicate that comorbidities do not predict clinical outcome in individuals having rotator cuff repair.

2.3.7 Involvement of the Dominant Arm:

Four studies examined whether a relationship existed between surgery on the dominant upper extremity versus the non-dominant upper extremity and the subsequent clinical outcome following rotator cuff repair. Two studies were prospective[4, 14], and two were retrospective studies.[15, 38] None of the 4 studies observed a relationship between the side of rotator cuff repair and postoperative clinical outcome scores.

2.3.8 Duration of Symptoms:

Six studies analyzed the relationship of duration of symptoms with postoperative clinical outcome scores. Bjornsson et al[31] and Petersen and Murphy[38], both retrospective studies, analyzed subjects with acute rotator cuff tears that were believed to be caused by trauma. They reported no relationship between clinical outcomes and the duration of time from onset of symptoms to surgical repair. In the remaining 4 studies[4, 10, 14, 16] comprised of individuals with chronic shoulder symptoms, duration of shoulder symptoms did not correlate with clinical outcome measures in 3 of the 4 studies. The 3 studies that did not demonstrate a relationship between duration of shoulder symptoms and outcome were prospective studies.[4, 14, 16] The 1 retrospective study found that a shorter duration of symptoms was correlated with better ASES and UCLA scores, but not with better Constant scores.[10] The studies of higher methodological quality indicate that duration of symptoms is not a strong predictor of clinical outcome following rotator cuff repair.

2.4 DISCUSSION

This systematic review of the literature revealed that the evidence for preoperative predictors of clinical outcome following rotator cuff repair were equivocal for most predictors. If any conclusion could be reached regarding the utility of the variables that were examined to predict clinical outcome of rotator cuff repair, the conclusion would be that the presence of a worker's compensation claim does result in worse clinical outcomes when measured by the DASH and that increasing age, while not being a consistent predictor of worse outcome, did predict worse outcomes on the DASH in the study with strongest methodology.

2.4.1 Discussion of predictors

In the studies examining the relationship of age to outcome, 6 of the 11 studies reported a negative relationship of age to clinical outcomes. However, in 2 of those 6 studies, the relationship of age to outcome was outcome-specific with not all of the clinical outcome measures used in the study demonstrating a significant correlation with age.[6, 14] For instance, in the study by Gartsman et al[6], age correlated to the Constant and UCLA, but not the ASES. In the study by Oh et al[14], there was a correlation of age to the SST, but not the Constant or ASES. The results of these two studies speak to the fact that relationships can differ with age based on the clinical outcome measure used.

Of the 5 studies that utilized MLR or logistic regression to examine age as an independent predictor of outcome, only 1 study, Milano et al[11], found age to remain an independent predictor of clinical outcome after other predictors of outcome were considered. However, this study was a randomized controlled trial that used a valid and reliable patient-reported outcome measure, the DASH, and it did receive a high score in terms of its quality (Table 2). Although there is a lack of consensus in the 11 studies, the fact that the strongest of these studies indicated that age is an independent predictor of worse clinical outcome, results in a weak conclusion that age is a negative predictor of the DASH score following rotator cuff repair. The American Academy of Orthopaedic Surgeons (AAOS), in a published 2010 guideline: *Optimizing the Management of Rotator Cuff Problems* also came to a "weak" recommendation for physicians to advise their patients that older age was correlated to less favorable outcomes following rotator cuff repair[48]. The AAOS recommendation that increasing age correlated with less favorable outcome was based on review of 23 studies published prior to October 2008. Twelve of these studies would not qualify for the current review due to the primary outcome being structural integrity of the tendon repair without reporting of a clinical, patient-reported outcome. The current review also includes 4 studies published after the October 2008 article deadline used in the AAOS guidelines. Also, due to different search strategies, there are 4 studies included in the AAOS guidelines that appear to meet the criteria for inclusion in the current review but were not included. While the composition of the studies included in AAOS guidelines and the current review differ to some degree, the study providing the strongest level of evidence indicated increasing age predicted less favorable DASH scores following rotator cuff repair. In addition, while some studies show no relationship between age and clinical outcome, there were no studies in this review that reported increasing age as an independent predictor of better clinical outcomes. Therefore, the strongest study indicated that age predicted worse DASH scores at 12 month follow-up and the remaining evidence does not support age as a predictor of outcome when measured by other outcome instruments.

Sex was examined as a predictor of outcome in 7 studies included in this review. Six of the 7 studies reported that there was no relationship between the sex of the patient undergoing rotator cuff repair and clinical outcome. Interestingly, only 1 of the studies used the DASH or WORC as the outcome used to assess the strength of sex as a predictor of outcome.[9] This study, by Henn et al, did not find sex to be a predictor of the DASH score. In addition to being the only study of the 7 which used the DASH as an outcome, it was also the strongest study in terms of quality (Table 2). Most of these seven studies used the Constant score and UCLA shoulder score as the outcome measure. Based on the current evidence, gender is not a consistent predictor of clinical outcome in individuals undergoing rotator cuff repair.

Fatty infiltration, as quantified by Goutallier et al[7], was examined for its prognostic ability in 6 studies that met the eligibility criteria for this systematic review (Table 5). The 3 prospective studies reported the GFDI was not a predictor of clinical outcome with the exception of the Milano et al study. The study by Milano et al[11], is a randomized trial, had the strongest quality score (Table 2) of the studies that examined the prognostic ability of GFDI scores and reported that a higher GDFI was an independent predictor of work-DASH scores but not the DASH or Constant score. The other 2 prospective studies, like the study by Milano et al, used multiple linear regression and found the GFDI was not an independent predictor of clinical outcome.[4, 14] Again, the outcome measures used differed between the studies with the Milano et al study using the DASH and the studies by Charousset et al and Oh et al using the Constant score as the outcome measure. In all, of the 6 studies that examined fatty infiltration, 4 reported no relationship of fatty infiltration to clinical outcomes, 1 reported GDFI to be a predictor of work-DASH but not the DASH or Constant score, and only 1, the study by Goutallier et al[7], reported a relationship between fatty infiltration and clinical outcome. In this case, the clinical outcome was the Constant score, an impairment-based outcome measure. A preponderance of evidence indicates that the GFDI is not a strong predictor of clinical outcome in individuals undergoing rotator cuff repair.

The effect of worker's compensation status on clinical outcomes was examined in 6 studies included in this systematic review (Table 6). All 6 of the studies received low quality scores (Table 2) indicating a need for stronger study designs and analyses to examine this predictor. Three studies indicated a worker's compensation claim was related to worse clinical outcomes[35-37] and 3 studies found no relationship of worker's compensation status to clinical outcome measures[4, 5, 38]. A closer analysis of Table 6 however, reveals that the two studies with the largest sample sizes and the largest numbers of worker's compensation claimants, Henn et al[35] and Namdari et al[37], found that a worker's compensation claim resulted in worse DASH score. Not only did these two studies have the largest samples, but they utilized the DASH, which is a reliable and validated outcome measure and used MLR to confirm that worker's compensation status was an independent predictor of outcome. The DASH is a patient-reported outcome measure and these results may indicate that individuals with a worker's compensation claim will report worse patient-reported outcomes following rotator cuff surgery. The 3 studies that found a worker's compensation claim to not predict outcome used the Constant, UCLA or ASES score; but not the DASH. Therefore, the results of this review agree with the 2010 AAOS guideline[48] which indicated there is moderate evidence to support advising patients that a worker's compensation claim correlates with less favorable outcome following rotator cuff repair. It also appears that a worker's compensation claim more readily predicts a patient-reported outcome but is less likely to predict an impairment-based outcome (Constant or UCLA score).

Although evaluated in three studies, the relationship between smoking and clinical outcome in patients undergoing rotator cuff repair was equivocal. The negative effects of smoking on collagen deposition and vascularity have made smoking a factor of interest as to its impact on rotator cuff healing following repair.[49] The retrospective nature of all 3 studies, the low quality scores (Table 2), as well as the lack of modern patient-reported outcome measures such as the DASH or WORC indicate the need for additional studies to investigate the effect of smoking on clinical outcome following rotator cuff repair.[15, 34, 38]

The two prospective studies that examined the relationship of comorbidities to clinical outcomes found no relationship between them.[3, 39] A retrospective study did report a relationship between the number of comorbidities and the postoperative DASH score.[37] While the stronger research designs and higher quality studies (Table 2) indicated no relationship between comorbidities and clinical outcomes, it must be taken into account that there were only 3 studies examining this relationship included in this review. Given the paucity of evidence on the effect of comorbidities on clinical outcomes following rotator cuff repair, no recommendation can be given regarding the prognostic ability of comorbidities although it seems that comorbidities do not affect patient-reported outcomes.

Duration of symptoms and involvement of the dominant arm as predictors of clinical outcome were evaluated in 6 and 4 studies, respectively. Neither of these factors demonstrated a relationship with clinical outcome in patients undergoing rotator cuff repair. 5 of the 6 studies that measured the relationship between duration of symptoms and clinical outcome found no relationship.[4, 10, 14, 16, 31, 38] For dominant arm involvement, all 4 studies reported no relationship among arm dominance and clinical outcome. [4, 14, 15, 38] However, all 4 of these studies used impairment-based outcomes and did not examine patient-reported outcomes. In addition all 4 studies received low quality scores based on the criteria in Table 2. Based on these studies, duration of symptoms does not predict clinical outcome following rotator cuff repair, but additional investigation of the effect of arm dominance on patient- reported clinical outcome is needed.

Potential predictors, reported in less than 3 research studies, were excluded from this review. This exclusion criterion was established to focus the review on predictors that have a "body of evidence" for their relationship to rotator cuff repair outcome. However, this exclusion criterion may have resulted in a strong predictor clinical outcome to be eliminated from the review. Therefore, all predictors, regardless of the number of times they were studied, were reviewed to see if any were found to be a strong predictor of outcome. Potential predictors such as BMI[37], work load[4], mechanism of cuff tearing[4], cuff atrophy[38], and patient expectations were excluded from this review. Of those predictors that were excluded, it appears that only patient expectations may predict outcome of rotator cuff repair.[9]

2.4.2 Discussion of the Outcome Measures

It is impossible to properly review the prognostic factors of clinical outcome after rotator cuff repair without an examination of the instruments used to measure outcome following surgery. A common, recurring theme in most studies included in this systematic review was a lack of a prospective study design, and a lack of modern, patient-centered, functional outcome measures. The Constant Shoulder Score, the American Shoulder and Elbow Surgeons Evaluation Form

(ASES), the UCLA Shoulder Score (UCLA), and the Simple Shoulder Test (SST) were commonly used to measure clinical outcome in this body of literature. However, all four of these outcome measures have significant limitations, which will be outlined in the following section, that call into question whether or not they should be used as primary outcome measures for studies investigating rotator cuff repair..[25] More modern, patient-centered measures of upper extremity function and quality of life, such as the DASH and the Western Ontario Rotator Cuff Index have not been used as frequently in this area of research. Modern outcome measures such as the Western Ontario Rotator Cuff Index are intended to measure the impact of rotator cuff disease on overall health.[20] The World Health Organization defines health as, "a state of complete physical, mental and social well-being." Previous outcome measures such as the Constant score, SST, UCLA score, and ASES cannot match the extent to which the WORC measures multiple dimensions of health and function.

Problems with the Constant score include a lack of information on its development as well as a lack of justification for the weighting used to calculate the final score.[25] The Constant score is heavily weighted towards impairment of shoulder motion and strength as they account for 65% of the final score. Questions regarding functional limitations experienced by the patient are limited to 3 yes/no questions. Due to the fact that the Constant score weighs impairment more heavily than function, this instrument is, in reality, an impairment-focused instrument rather than a patient-focused instrument that measures activity limitations and participation restrictions. In a 2009 systematic review by Roy et al[50], the Constant was excluded from a review of its psychometric properties due to the lack of studies that have evaluated these properties.

The ASES was designed by a committee of experts with no input from patients to generate or select items.[25] Fifty percent of the final score is determined by the patient's response on a visual analog scale that determines their pain level. As with the Constant score, the weighting of the various items is not explained. The ASES also uses a 4-point Likert scale to query a patient regarding their limitations in 10 activities that involve the shoulder. The 4-point scale uses a "2" to indicate "somewhat difficult" and a "3" to indicate "not difficult". A patient that initially indicates the activity is "somewhat difficult" needs to improve to a level of "not difficult" to demonstrate any improvement in function. Due to this fact, the ability of the ASES to be responsive has been called into question.[25, 50]

The UCLA shoulder score, although developed for use in evaluating the outcome of shoulder arthroplasty[51], has become a frequent outcome measure in studies of rotator cuff pathology. This instrument lacks a detailed explanation of its development.[52] The UCLA is comprised of only 5 questions, one each pertaining to pain, function, active shoulder flexion, strength of shoulder flexion, and patient satisfaction. Like the Constant Score and ASES, the UCLA score also has arbitrary weighting of each question. The heavy weighting toward measures of impairments makes the UCLA an impairment-focused instrument rather than a patient-centered instrument. Finally, because the 5th question asks if the patient is satisfied with the outcome, this instrument may not be used in clinical trials where a baseline measurement of shoulder function is desired.

The Simple Shoulder Test (SST) contains 12 questions that are answered yes/no in an attempt to quantify the activity limitations of the patient. A justification of why these 12 questions were chosen is not provided by the instrument's developers.[52] Like the ASES, which utilizes a Likert scale for its responses, the limited response choices available on the SST theoretically should limit its responsiveness. While the SST has been shown to be reliable, Roy et al[50] also called into question the instrument's responsiveness given its dichotomous response scale. Somewhat surprisingly, the SST has demonstrated acceptable responsiveness in past studies involving patients undergoing rotator cuff repair.[53, 54]

Of the clinical outcome measures used in the studies that were included in this systematic review, the WORC and the DASH have been developed using the most rigorous methodologies.[20, 24, 25] These methodologies include involvement of actual patients in the item development and thorough testing of the items within the instrument to determine the most parsimonious set of questions for the instrument. In addition, the WORC and DASH attempt to include additional domains of health such as mental and social well-being as defined by the World Health Organization.[55] In previous studies, both the WORC and the DASH have been shown to have good psychometric properties.[23, 27]

2.5 LIMITATIONS

There are limitations to this systematic review and the results that are drawn from a review of this particular body of literature. Publication bias may lead to more frequent publication of data when potential prognostic factors are shown to have a significant relationship with postoperative clinical outcome measures. It was difficult to locate prognostic studies whose primary purpose was to investigate preoperative factors that predict the outcome rotator cuff repair. Many of the studies that did look at prognostic factors were outcomes studies that considered prognosis as as a secondary analysis. Therefore, many studies were not retrieved by searching Medline or Embase using the MeSH term “prognosis”. Another problem is that the relationships found between predictors and clinical outcome measures often conflicted within the same study. This indicates that the relationship is specific to the outcome measure. The lack of a “gold standard” instrument for assessing clinical outcome after rotator cuff repair makes it extremely difficult to compare findings between studies as each study uses a different outcome measure. In addition to the lack of a gold standard outcome measure for rotator cuff repairs, commonly used outcome measures such as the Constant Shoulder Score and the UCLA Score lack appropriate development and psychometric testing. Finally, most studies included in this systematic review make no attempt to control or record concomitant procedures that occur during rotator cuff repair. For instance, procedures such as labral debridement, SLAP repair, subacromial decompression, biceps tenodesis or tenotomy need to be recorded and controlled for during analyses of predictors to determine their impact on clinical outcome following rotator cuff repair.

2.6 FUTURE RECOMMENDATIONS

1. Since the prognostic ability of a preoperative predictor of outcome will vary depending on the outcome measure (i.e. ASES, SST, DASH), a gold standard outcome measure should be used by all prognostic studies investigating predictors of rotator cuff repair. Implementing a gold standard outcome measure will allow more useful comparisons among studies.
2. Modern outcome instruments; developed with sound methodological processes, patient input with regards to content, adequate psychometric properties, and the ability to measure the multi-dimensional aspects of health, should replace older instruments that lack these properties.
3. Prognostic studies that examine modifiable impairments such as deficits in glenohumeral motion, rotator cuff strength and scapular dysfunction should be conducted to determine

if modifiable impairments, that are amenable to treatment, can affect rotator cuff outcome. Such modifiable factors may also include psychosocial factors such as fear-avoidance, anxiety, or depression.

4. Advances in imaging will allow better accuracy and reliability in determining the preoperative size and location of a cuff tear. Since many studies have measured the intraoperative size of cuff tears and shown its relation to the integrity of the cuff repair, future prognostic studies will be able to preoperatively measure this factor and use that information to predict clinical outcome of rotator cuff repair.

5. Future studies should be prospective rather than retrospective and utilize multivariable methods to analyze the influence of multiple potential predictors.

2.7 CONCLUSION

The limitations in study design and methodology as well as the often conflicting prognostic evidence precludes definitive conclusions regarding the prognostic ability of the reviewed factors.[56] At best, it appears the presence of a worker's compensation claim and possibly increasing age have negative impacts on clinical outcome following rotator cuff repair.

2.8 FIGURES

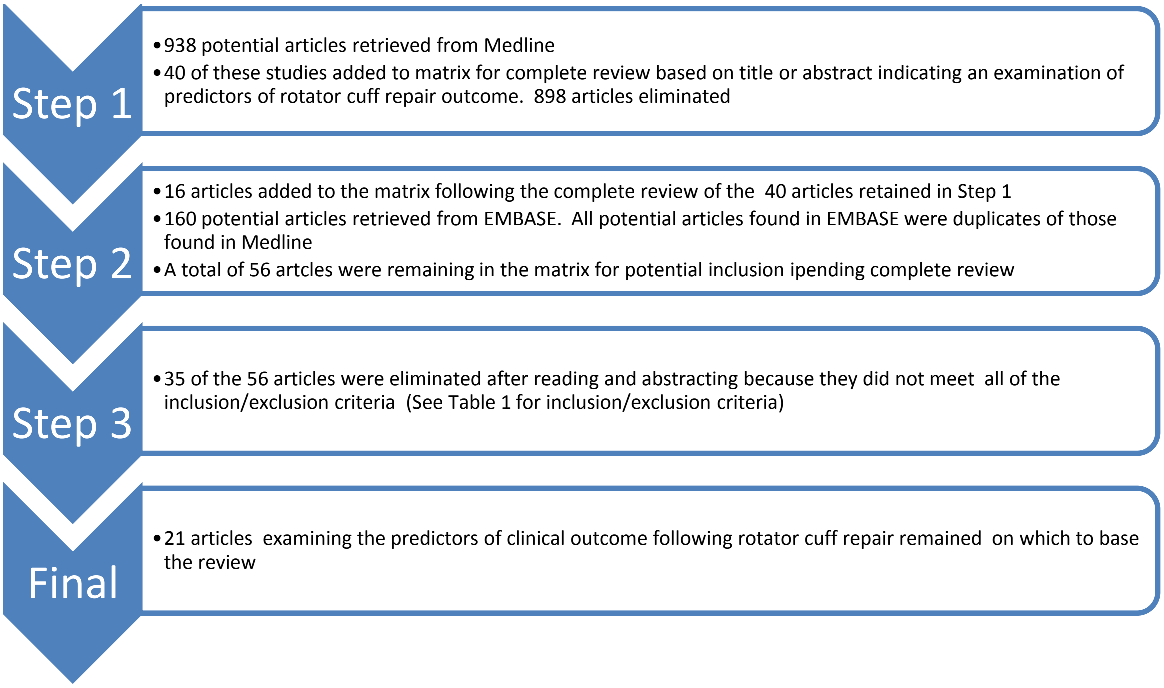


Figure 1 Process of Article Selection

2.9 TABLES

Table 1 Final inclusion criteria for articles in the review

Publication Details	***Included****:* Peer-reviewed articles published 1995 or later. ***Excluded:*** 1. Articles not available in English 2. Studies published 1995 to present, but analyzing data collected prior to 1990. This was determined by recording when the surgeries were performed. 3. Articles specifically focusing on repair of massive rotator cuff tears
Study Design Types	***Included****:* All Designs
Population Characteristics	Individuals with a full thickness cuff tear that underwent open or arthroscopic repair.
Title and Abstract Requirements	Articles selected had to indicate in either the title or the abstract that they analyzed a predictor of outcome of rotator cuff repair.
Outcome Measures	The study must report at least 1 disability/function/quality of life outcome instrument. Articles examining only the predictors of structural integrity of the rotator cuff repair were excluded.
Selection of Predictors	To be included in the review, the predictor being examined for its relationship to clinical outcome must have been investigated in at least 3 separate studies.

Table 2 Summary of Study Quality

		DESIGN		ATTRITION	VALIDATION	STATISTICAL ANALYSES			OUTCOME TYPE		QUALITY SCORE*
Author	Year	Prospective study design	Multi-site study	Follow-up was ≥ 80% of total number of subjects enrolled at baseline	Were the results validated in a 2nd independent sample	Strength of relationship between predictor and outcome variable reported	Strength of predictor reported after controlling for confounding/ competing predictor variables	Assessment of Model Strength (ie. R^2 or Pseudo R^2/Area under curve)	Impairment -based outcome measure	Patient-reported outcome measure	
Boissonault	2007	Y	Y	N	N	Y	Y	Y	none	DASH	5
Milano	2007	Y	N	Y	N	Y	Y	N	none	DASH	4
Henn	2007	Y	N	Y	N	Y	Y	N	none	SST DASH	4
Gulotta	2011	Y	N	N	N	Y	Y	N	none	ASES	3
Tashjian	2006	Y	N	N	N	Y	Y	N	none	DASH SST	3
Gartsman	1998	Y	N	Y	N	Y	N	N	Constant UCLA	ASES	3
Namdari	2010	N	N	NA	N	Y	Y	N	none	DASH	2
Oh	2009	Y	N	N	N	Y	N	N	Constant	SST ASES	2
Charousset	2008	Y	N	Y	N	N	N	N	Constant	none	2
Henn	2008	Y	N	Y	N	N	N	N	none	DASH SST	2
Liem	2007	Y	N	Y	N	N	N	N	Constant	none	2
Baysal	2005	Y	N	Y	N	N	N	N	none	ASES, WORC	2
Prasad	2005	Y	N	Y	N	N	N	N	Constant	none	2
Mallon	2004	N	N	NA	N	Y	Y	N	UCLA	none	2
Pai	2001	Y	N	Y	N	N	N	N	Constant UCLA	none	2
McKee	2000	Y	N	Y	N	N	N	N	none	SPADI	2
Cole	2007	Y	N	N	N	N	N	N	Constant	ASES SST, ROWE	1
Goutallier	2003	N	Y	NA	N	N	N	N	Constant	none	1
Bjornsson	2011	N	N	NA	N	N	N	N	Constant	DASH WORC	0
Petersen	2011	N	N	NA	N	N	N	N	UCLA	ASES	0
Oh	2010	N	N	NA	N	N	N	N	Constant UCLA	ASES SST	0
Hersch	2000	N	N	NA	N	N	N	N	UCLA	ASES	0

Legend: Y, yes; N, no. *Maximum score of 7 “Yes” responses with higher scores indicating stronger study design and analysis.

Table 3 Age as a predictor of clinical outcome

Author, year	Study Design/ Quality Score	N	Time to follow-up	Outcome Measure	Method to assess relationship	Correlated or Predictive	No association
Milano et al 2007	PRC/4	80	24 mo.	DASH	Univariate and multivariable regression	Age was an independent predictor of a worse DASH score in a multivariate regression model	
Gulotta et al 2011	PC/3	106	60mo	ASES	Logistic regression		Age was not a significant predictor of an ASES score greater than 90
Gartsman GM et al 1998	PC/3	50	13 mo.	UCLA, ASES, Constant	Correlation of age to UCLA r=.39; age to Constant r=.30; age to ASES r=.12	Age had a significant negative correlation to the Constant and UCLA score but not he ASES	
Oh et al 2009	PC/2	78	20mo	Constant, SST, and ASES	Correlations and stepwise linear regression	Correlation of SST to age (r= -.30).	It does not appear that age was an independent predictor of outcome in the regression models. Constant and ASES not correlated to age

Table 3 (continued)

Author, year	Study Design/ Quality Score	N	Time to follow-up	Outcome Measure	Method to assess relationship	Correlated or Predictive	No association
Liem et al 2007	PC/2	53	26mo	Constant	Table I indicates a p-value of .002 for the influence of age on the Constant score. This appears to be a correlation of age to the Constant score but this is not explicitly stated.	Higher age was associated with lower Constant scores	
Pai et al 2001	PC/2	54	34 mo.	Constant and UCLA	Chi-squared compared rates of excellent, and good to fair or poor outcomes based on the UCLA by decade of age		No correlation of age to outcome
Baysal et al 2005	PC/2	60	12mo.	ASES, WORC	ANOVA to compare ASES and WORC scores of <50 yrs, 50-59 yrs, and ≥60		Age did not influence postoperative ASES or WORC scores
Prasad et al 2005	PC/2	42	26 mo.	Constant	Correlation	Older age was significantly correlated to worse Constant scores	
Bjornsson et al 2011	R/0	42	39mo	Constant, DASH, WORC	t-tests to compare functional outcome scores of those older to younger than 65		No difference in functional outcome scores in patients younger versus older than 65

Table 3 (continued)

Author, year	Study Design/ Quality Score	N	Time to follow-up	Outcome Measure	Method to assess relationship	Correlated or Predictive	No association
Oh et al 2010	R/0	177	29mo	Constant, ASES, SST, UCLA	Correlation and MLR	Age positively correlated to Constant score (magnitude not reported)	Although age correlated to a better Constant score, it was not an independent predictor of functional outcome in final model
Hersch et al 2000	R/0	22	39 mo.	Constant, ASES, UCLA	Correlation and Multivariable regression. Neither the strengths of the associations or the regression models are reported	"older age was significantly associated with a better outcome only for the ASES and UCLA scoring systems"	Age was not a predictor in multivariable regression analysis

Abbreviations: PC= prospective cohort; R= retrospective; PRC=Prospective Randomized controlled trial; MLR=multiple linear regression; mo=months; ASES=American Shoulder and Elbow Surgeons Score; SST=Simple Shoulder Test

Table 4 Sex as a predictor of clinical outcome

Author, year	Study Design/ Quality Score	N	Time to follow-up	Outcome Measure	Method to assess relationship	Correlated or Predictive	No association
Henn RF et al, 2007	PC/4	125	12 mo.	SST, DASH	Multiple linear Regression		Sex was not a predictor of any outcome in the study
Gartsman GM et al 1998	PC/3	50	13 mo.	UCLA, ASES, Constant	Wilcoxon rank-sum test		Women did not have significantly different Constant, ASES or UCLA scores than men
Oh JH et al, 2009	PC/2	78	19 mo.	SST, Constant and ASES	Correlation (r=.29, p=.038) and multiple linear regression	Being female correlated to worse SST scores but not to Constant or ASES scores	No mention of the results of the regression analysis
Charousset C et al, 2008	PC/2	114	31 mo.	Constant	Multiple linear regression	Being female was a significant negative predictor of the Constant score	
Mallon WJ et al, 2004	R/2	224	≥ 12 mo. (Mean not reported)	UCLA	t-test		Women did not have UCLA scores that significantly differed from men
Petersen and Murphy, 2011	R/0	36	31 mo.	ASES score, UCLA	t-test		Clinical outcomes did not differ between men and women
Oh JH et al, 2010	R/0	177	29 mo.	Constant, ASES, SST UCLA	Correlation		Sex was not correlated with the clinical outcomes

Abbreviations: PC= prospective cohort; R= retrospective; PRC=Prospective Randomized controlled trial; MLR=multiple linear regression; mo=months; ASES=American Shoulder and Elbow Surgeons Score; SST=Simple Shoulder Test; UCLA=University of California Los Angeles Shoulder Score

Table 5 Fatty infiltration of the rotator cuff musculature as a predictor of clinical outcome

Author, year	Study Design/ Quality Score	N	Time to follow-up	Outcome Measure	Method to assess relationship	Correlated or Predictive	No association
Milano et al 2007	PRC/4	80	24 mo.	Constant, DASH, Work-DASH	Univariate and multivariate regression	A higher GFDI score was an independent predictor of the Work-DASH	GFDI not associated with DASH or Constant score
Oh J, 2009	PC/2	78	19 mo.	Constant, SST, ASES	Univariate and MLR		Fatty degeneration was not related to clinical outcomes
Charousset C, 2008	PC/2	114	31 mo.	Constant	Multiple linear regression		GFDI was not associated with Constant score
Goutallier D, 2003	R/1	220	37 mo	Constant score	Correlation (not specifically stated)	A higher presurgical GFDI score was associated with a worse Constant score	
Petersen and Murphy, 2011	R/0	36	31 mo	ASES, UCLA	t-test		No difference in clinical outcomes between various levels of GFDI
Oh et al 2010	R/0	177	29mo	Constant, ASES, SST, UCLA	1.correlations 2. MLR		Fatty degeneration was not related to clinical outcomes

Abbreviations: PC= prospective cohort; R= retrospective; PRC=Prospective Randomized controlled trial; MLR=multiple linear regression; mo=months; ASES=American Shoulder and Elbow Surgeons Score; SST=Simple Shoulder Test; UCLA=University of California Los Angeles Shoulder Score

Table 6 Workers compensation claim as a predictor of clinical outcome

Author, year	Study Design/ Quality Score	N	Time to follow-up	Outcome Measure	Method to assess relationship	Correlated or Predictive	No association
Henn et al, 2008	PC/4	N=125 (39 with WC claim)	12 mo.	DASH, SST	1.MLR with the DASH as the dependent variable 2. t-tests	Patients with a WC claim had statistically worse DASH and SST outcome scores. Having a WC claim remained an independent predictor of the 12 month DASH score	
Namdari et al, 2010	R/2	154 (43 with WC claim)	12 mo.	DASH	MLR	Worse DASH scores were associated with having a WC claim. WC status was an independent predictor of worse DASH score.	WC status not a predictor of change in DASH scores
Charouss et et al, 2008	PC/2	114 (12 with work-related injury)	31 mo.	Constant	MLR		A work-related shoulder injury did not predict clinical outcome
McKee et al, 2000	PC/2	67 (23 with WC claim)	24 mo.	SPADI	t-test SPADI 81 for No WC claim and 62 for WC claim (p=.01)	Patients with WC claims had lower SPADI scores (a lower score indicates a worse outcome)than those patients without a WC claim	
Cole et al, 2007	PC/1	49 (22 with WC claim)	32 mo.	Constant, ASES, SST, Rowe score	Mann-Whitney U, χ^2		Having a WC claim did not result in worse clinical outcome.
Petersen and Murphy, 2011	R/0	36 (12 with WC claim)	31 months	ASES, UCLA	t-test		No difference in clinical outcomes between WC and Non-WC

Abbreviations: PC= prospective cohort; R= retrospective; PRC=Prospective Randomized controlled trial; MLR=multiple linear regression; mo=months; ASES=American Shoulder and Elbow Surgeons Score; SPADI=Shoulder Pain and Disability Index; SST=Simple Shoulder Test; UCLA=University of California Los Angeles Shoulder Score

3.0 PREDICTORS OF SHORT-TERM PATIENT-REPORTED OUTCOME FOLLOWING SURGICAL TREATMENT OF ROTATOR CUFF PATHOLOGY

3.1 INTRODUCTION

Given the increased prevalence of rotator cuff related surgeries over the last decade[29, 48], and the 2008 report published by the American Academy of Orthopedic Surgeons which forecasts a growing number of rotator cuff lesions for the aging American population[57], there will be an increasing number of stakeholders interested in prognosis for functional recovery in individuals considering rotator cuff surgery. Providing an evidence-based prognosis for the outcome of rotator cuff surgery should be a goal of those involved in caring for these individuals. Past studies have shown evidence-based models to be superior to subjective judgments of the provider in predicting health outcomes.[58, 59] Accuracy in predicting an individual's patient-reported outcome (PRO) for rotator cuff surgery is important because it can help set realistic expectations regarding outcome. In addition, prognoses determined from an evidence-based model could aid individuals in their decision to undergo elective shoulder surgery.

Predictors that have been frequently investigated for their relationship to clinical outcome in individuals undergoing surgery for rotator cuff pathology include age[2, 6, 10, 11, 14-16, 31, 33, 60], sex of the patient[4, 6, 9, 14, 15, 34, 38], fatty infiltration of the rotator cuff muscles[4, 7, 11, 14, 15, 38], worker's compensation status[4, 5, 35-38], smoking status[15, 34, 38], comorbidities[3, 37, 39], involvement of the dominant arm[4, 14, 15], and duration of symptoms.[4, 10, 14, 16] Of these predictors, having a worker's compensation claim has the strongest evidence to support its negative effect on PRO. Increasing age has somewhat more equivocal evidence, but the study by Milano et al[11], a randomized trial which used regression analysis to control for potential confounding predictors, found increasing age to be an independent predictor of worse DASH scores following rotator cuff repair with or without subacromial decompression. Interestingly, with the exceptions of worker's compensation and age, the remaining prognostic factors have demonstrated little prognostic ability to predict patient-reported outcome after rotator cuff repair.

While past studies have examined predictors of outcome in individuals undergoing rotator cuff repair and subacromial decompression, a majority of the studies were not prospective in design and do not utilize regression modeling to determine the strength of a predictor while considering the influence of other potential predictors. In addition, the clinical outcome used in many of these prior studies is often lacking in terms of its psychometric properties. For instance, clinical outcome instruments, such as the Constant Shoulder Score and the UCLA Shoulder Score, are frequently used as the clinical outcome instruments in these studies in spite of their lack of psychometric testing and lack of components that capture the multi-dimensional nature of health as defined by the World Health Organization.[52] In addition, these instruments both contain items that are scored by the clinician, such as shoulder strength and range of motion. As such, these outcome measures are not patient-centered and are susceptible to rater bias.

In addition to the above concerns regarding many of the existing prognostic studies, there are also limited prior studies investigating the prognostic value of impaired rotator cuff strength, glenohumeral internal rotation, scapular dysfunction, shoulder activity levels, and psychosocial variables. Measurements of shoulder strength and motion, as well as observation of scapular motion for evidence of dyskinesis are ubiquitous in the examination of an individual with rotator cuff pathology. Deficits in internal rotation range of motion[61-63] , external rotation strength[64, 65], and scapular dyskinesis in individuals with rotator cuff pathology have been demonstrated.[63, 66, 67] Given these past findings, the severity or presence of these impairments prior to surgery may predict clinical outcome after rotator cuff surgery and should be investigated further.

Finally, the impact of psychosocial factors on patient-reported outcome of rotator cuff surgery has not been investigated for its prognostic ability in individuals with rotator cuff pathology who undergo surgical intervention. Fear-avoidance behaviors, depression and anxiety have been shown to be predictors of outcome in the treatment of individuals with low back pain and arthritic conditions. [68 69 70, 71] but have not been investigated in conditions involving rotator cuff pathology. George et al did demonstrate that fear of pain and anxiety levels were predictive of one's DASH score 24 hours following an exercise protocol designed to induce muscle soreness.[72] However, this study was performed on healthy individuals and only served to provide preliminary evidence that fear-avoidance behaviors and anxiety could impact clinical outcomes in individuals with shoulder pathology. As for depression as a predictor of outcome, Ring et al[73] have shown the Center for Epidemiologic Studies Depression Scale score (CES-D) to be an independent predictor of the DASH score in patients with various upper extremity disorders. Given the relationship of psychosocial factors to PRO for other patient populations, investigation of the prognostic ability of psychosocial factors should be investigated for individuals undergoing surgical treatment of rotator cuff pathology.

The purpose of this study was to utilize preoperative, clinically-derived data to identify predictors of PRO 6-months following surgery for rotator cuff pathology. The prognostic factors that were explored in this study included: patient characteristics, history of the shoulder injury, symptoms, impairment of glenohumeral motion and strength; scapular dyskinesis, preoperative disability as measured by WORC and DASH scores, shoulder activity level, and psychosocial measures of fear-avoidance, depression, and anxiety. To provide a strong level of evidence, this study utilized a prospective cohort design, logistic regression analysis, and a valid, reliable and responsive PRO instrument, the Western Ontario Rotator Cuff Index, which is a measure capable of capturing the multi-dimensional nature of health.[20]

3.2 METHODS

3.2.1 Subjects

From November 2008 to May 2011, approximately 85 patients were approached regarding participation in this study. For those choosing not to participate, the reason for non-participation was lack of time to complete the baseline testing necessary for the subject to participate in this study. A total of 65 patients from the University of Pittsburgh Medical Center (UPMC) Center for Sports Medicine agreed to participate and were enrolled in this study. These individuals were scheduled to undergo arthroscopic subacromial decompression with or without arthroscopic repair of a small rotator cuff tear. All surgeries were performed by a single surgeon (MR). The Institutional Review Board at the University of Pittsburgh approved the study, all patients provided informed consent, and the rights of the individuals enrolled as subjects were protected at all times.

Three individuals who originally planned to have surgery and provided consent for participation in the study subsequently did not undergo surgery but instead chose to pursue continued conservative treatment and were excluded from the study. Therefore 62 patients were included in this prospective cohort based on the following eligibility criteria. Patients scheduled for arthroscopic subacromial decompression with or without a partial or full thickness rotator cuff tear that was no greater than 2cm in diameter were enrolled in the study. In addition, to participate in the study, individuals had to be 25 years of age or older and have a primary complaint of shoulder pain for at least 3 months, a positive Kennedy-Hawkins or Neer's impingement test and a pain-free contralateral shoulder. The exclusion criteria were: 1) prior glenohumeral dislocation that required relocation by a physician; 2) prior shoulder surgery on the affected side; 3) rheumatoid arthritis; 4) active cervical radiculopathy or 5) expected Grade II or worse labral tear that would require repair at the time of surgery. Those patients with an unexpected labral tear, discovered during the arthroscopic surgery, were not excluded from the study.

3.2.2 Data Collection

Patients that agreed to participate in this study participated in a single preoperative data collection session. Data collection included demographic information (age, gender, level of education, employment status, and presence of worker's compensation claim), history of the present shoulder condition, and social history. Additional surveys were used to measure the patient's shoulder activity level, fear-avoidance, depression, and anxiety levels (described below). Impairments in the range of glenohumeral internal rotation, rotator cuff strength, and scapular dyskinesis were also measured as described below. Patient-reported outcome measures included a disease-specific PRO, the Western Ontario Rotator Cuff Index (WORC), a region-specific patient-reported outcome instrument, the Disabilities of the Arm, Shoulder and Hand

Questionnaire (DASH), and a global rating of change. The 3 PRO measures were also collected via mailed survey 6 months after surgery and served as the primary endpoint for analyses.

3.2.2.1 Shoulder Activity Level

To assess the daily demands placed on the shoulder, the Shoulder Activity Scale, developed by Brophy et al[74] was administered at baseline. The scale ranges in value from 0 to 20 with 0 being no activity involving the shoulder and 20 being maximum shoulder activity. The scale asks 5 questions. The questions are: how often do you: 1) carry objects greater than 8 pounds (such as a bag of groceries); 2) handle objects overhead; 3) weightlift with the arms; 4) perform swinging motions (such as in tennis, golf, baseball); and 5) lift items heavier than 25 pounds (not including weightlifting). The possible responses are: 1 - never or less than once a month (0 points), 2 - once a month (1 point), 3 - once a week (2 points), 4 - more than once a week (3 points) or 5 - daily (4 points). The scale has been shown to be reliable.[74]

3.2.2.2 Psychosocial Measures

The Fear Avoidance Beliefs Questionnaire (FABQ) was used to quantify fear-avoidance beliefs (See Appendix, Table 25, for a full copy of the FABQ). The FABQ consists of 16 statements and attempts to quantify how much fear -avoidance behavior a person is experiencing. These statements generally focus on whether or not the respondent believes that physical activity of the shoulder will worsen their condition. It also focuses on whether or not the respondent believes their work is too difficult for their current shoulder condition and if they believe they will be able to continue their current work now and in the future. A 7-point Likert scale is used to score the individual's agreement with statement as: 0 - "completely disagree" with the statement, 3 - "unsure" about the statement, 6 - "completely agree" with the statement. The full FABQ score includes all 16 items. However, there are two subscales that can be calculated. Seven items comprise the work subscale with a maximum score of 42. Four items comprise the physical activity subscale with a maximum score of 24 on this subscale. Therefore, there are 5 items that are part of the total score, but not part of the two subscales. The FABQ has been shown to have good reliability and internal consistency in patients with various musculoskeletal diagnoses including low back pain and shoulder pain.[75-77] To make the FABQ applicable for subjects with shoulder pathology, the instructions were changed to query the subject on how physical activity affects their shoulder pain rather than how physical activity affects their low back pain.

The CES-D, a 20 item self-administered questionnaire, was used to quantify depressive symptoms. Subjects were asked how frequently each of the 20 symptoms occurred over the past week. The CES-D uses a 4-point Likert scale with 0 -"rarely or none of the time" (less than 1 day), 1 -"some or a little of the time" (1-2 days), 2 -"occasionally or a moderate amount of time" (3-4 days), 3 -"most or all of the time" (5-7 days). Scores can range from 0-60 with higher scores indicating more depressive symptomatology. A score of 16 or greater is used as the cut-point that may indicate clinical depression. Initial test-retest reliability of this measure was found to have a correlation of 0.59 at 8 weeks.[78] More recently, Hann et al[79] found the CES-D to have ICCs of 0.87 and 0.89 at a mean test-retest interval of 2.5 weeks in a healthy control group and a group of women undergoing treatment for breast cancer, respectively.

The Beck Anxiety Inventory (BAI) was used to measure the symptoms of clinical anxiety. The inventory consists of 21 questions and uses a 4-point Likert scale. Respondents are asked how much they were bothered by each symptom over the past week including today. The responses are: 0 -"Not at all"; 1 -"Mildly" (it did not bother me much); 2 -"Moderately" (it was very unpleasant but I could stand it"; 3 -"Severely" (I could barely stand it). The highest level of anxiety as measured by the BAI is 63 and the lowest score possible is 0. A score of 10-18 points is indicative of mild anxiety while scores above 19 are indicative of moderate anxiety[80]. Beck et al reported the test-retest reliability to be 0.75 in a group of adult outpatients with various psychiatric diagnoses while Osman et al found similar test-retest reliability of 0.71 in a population of adolescents with psychiatric diagnoses.[81, 82]

3.2.2.3 Measures of Shoulder Impairment

A bubble goniometer was used to record the range of active internal rotation motion the patient was able to achieve. The scapula was stabilized with a posteriorly directed force through the coracoid and lateral aspect of the scapula to prevent anterior tipping of the scapula. Previous studies have reported intra-class correlation coefficients of .62 to .71 for the intra-tester reliability of this measurement.[83]

Strength of the infraspinatus and supraspinatus was measured using a Lafayette handheld dynamometer (Model 01163. Lafayette, IN) with the techniques described and validated by Kelly et al.[84] Acceptable reliability of using a handheld dynamometer for assessing strength of these rotator cuff muscles has been reported.[85] However, testing of the subscapularis, in the position validated by Kelly et al, was not possible due to increased pain reported by subjects with the hand behind the back in the “lift-off” position. Therefore, internal rotation strength was measured with the arm at the side, in neutral glenohumeral rotation, and the elbow at 90° of flexion. This position of testing internal rotation strength has been shown reliable but lacks EMG validation.[85]

Three measures of scapular dysfunction, the Scapular Assist Test (SAT), the Scapular Dyskinesis System (SDS) and the Scapular Index score were used to quantify scapular dysfunction. Rabin et al established the SAT, which was originally described by Kibler, to be a reliable test.[86] During the SAT, the physical therapist assists the upward rotation and posterior tilting of the scapula while the patient actively elevates the affected arm.[86, 87] The test is considered positive if assistance with scapular motion during elevation of the arm decreases pain greater than 2 points on an 11-point numerical pain rating scale as compared to elevation of the arm without scapular assistance. The Scapular Dyskinesis System allows for the qualitative classification of scapular position at rest and during elevation of the arm.[88] The system categorizes patients with shoulder pathology into 1 of 4 categories based on the position of the scapula: 1) prominent inferior angle; 2) prominent medial border; 3) elevated superior border and 4) symmetric scapula. Moderate inter-rater reliability has been established for this measurement system.[88] The Scapula Index provides a measure of the influence of the pectoralis minor on the resting position of the scapula.[89] To calculate the Scapula Index, the examiner measures from the sternal notch to the medial aspect of the coracoid process. The examiner then measures the horizontal distance from the posterolateral corner of the acromion to the thoracic spine. The sternal notch to coracoid process distance is then divided by the distance from the posterolateral corner of the acromion to the thoracic spine. The resulting value is multiplied by 100. (Photos depicting the measures of scapular dysfunction and scapular positioning are included in the Appendix.)

3.2.2.4 Operative Data

The location and degree (specific cuff tendon and no tear, partial tear, full tear) of the rotator cuff were recorded intra-operatively. Due to the many possible combinations of tear severity and rotator cuff tendon(s) involved, the rotator cuff tear pattern was collapsed into one of three possible groups: 1) no cuff tear 2) partial cuff tear 3) full cuff tear. The surgical procedures performed may also result in a large number of repair combinations, therefore the surgical procedures were collapsed into 1 of the following 4 groups: 1) subacromial decompression only 2) subacromial decompression and repair of the rotator cuff tendon 3) subacromial decompression, rotator cuff tendon repair and either labral repair or biceps tenodesis 4) subacromial decompression and labral repair with or without biceps tenodesis. From these data, the potential confounding effects of the severity of rotator cuff damage and surgical procedure on PRO were assessed.

3.2.2.5 PRO Measures

Both the WORC and the DASH were assessed preoperatively and 6-months after surgery. The disease-specific WORC is scored from 0 to 100 with higher scores indicating better health and function. The WORC includes 21 items that are each answered on a 100 mm visual analog scale with anchors that include no pain/difficulty and extreme pain/difficulty. For each item, the patient can receive a score that ranges from zero (no problem or loss of ability) to 100 (maximum pain/difficulty/disability). The scores for each item are summed and transformed to a scale that ranges from 0 to 100 by subtracting the raw score from 2100, dividing by 2100 and multiplying by 100. The reliability and responsiveness of the WORC has been established in earlier studies[20, 27]

The DASH is an upper extremity, region-specific instrument comprised of 30 items with each item answered on a 5-point likert scale. Typical response choices for each item range from "no difficulty" performing the task/activity to "unable" to perform the task or activity. Total scores range from 0 to 100 with higher scores indicating worse health and function.[5, 24] The DASH has been shown to have strong psychometric properties.[23]

The global rating of change was also measured at the 6-month postoperative time point. The global rating of change is a 15-point scale that ranges from -7 (a very great deal worse) to +7 (a very great deal better) with 0 in the middle (about the same). Scores of +4 and +5 are indicative of moderate changes in status while +6 and +7 are indicative of a large change in a person's status.[90] The GROC has been supported for use as an outcome that measures the patient's perspective as to whether or not they are better.[91]

3.2.3 Criteria for determining "Responders" and "Nonresponders"

Individuals who responded most positively to surgical intervention were classified as "responders" while those who demonstrated less improvement with surgical intervention were classified as "nonresponders." The 6-month postoperative WORC change score and the global rating of change were the criteria used to classify each individual as a responder or nonresponder. To be considered a responder, the patient had to report at least a 17 point improvement from their baseline WORC score as well as report a score of at least +5 or better on the global rating of change ("quite a bit better"). It can be argued that these criteria for determining a responder are more stringent than those often used in this area of research in that the patient had to not only report a high level of improvement on the global rating of change, they also needed to report substantial functional improvement as evidenced by their improved WORC score. Previously, Kirkley et al [20] have shown a 17-point change on the WORC to be the minimal clinically important change in individuals undergoing cortisone injections for subacromial impingement syndrome. We are currently unaware of any study determining the minimal clinically important change in individuals undergoing arthroscopic subacromial decompression with or without a rotator cuff repair.

3.2.4 Statistical Analysis

Power analysis determined a sample of 50 individuals would provide 80% power for logistic regression analysis to detect an improvement in predictive accuracy from 68% to 85%.[92] Predictive accuracy of 68% was based on a previous randomized controlled trial that reported a success rate of 68% in individuals who underwent arthroscopic subacromial decompression for chronic subacromial impingement syndrome.[93] If the clinician always predicted success in individuals having this surgery, they would be correct 68% of the time without any additional information. The goal of this study was to construct a model-based estimate of being a responder or nonresponder that was highly accurate. Choosing a target accuracy of 85% for our model allowed for high discriminative accuracy while maintaining a feasible sample size. Descriptive statistics were calculated for all potential prognostic factors measured at baseline as well as for the PRO measures at baseline and 6-month follow-up.

3.2.4.1 Logistic Model Building Process

Univariate logistic regression was used to determine if the odds ratio for each potential prognostic variable was a statistically significant predictor of outcome at the $p<0.05$ level. The potential predictors that had a statistically significant univariate relationship with the outcome were examined in a direct entry multivariable logistic regression models as well as in backwards elimination models.

In the direct entry models, non-modifiable factors were entered prior to inclusion of modifiable factors. This was done due to the fact that non-modifiable factors cannot be changed, so they need to be assessed and accounted for in the model prior to examining the predictive ability of factors that are potentially modifiable. The non-modifiable potential predictor variables were entered into the model based on the p-value from univariate analysis. The factor with the smallest p-value was entered first and then remaining non-modifiable factors were entered sequentially based on their univariate analysis p-value.

Once the non-modifiable factors were entered into the model, the potentially modifiable factors were entered using the same rules applied to the non-modifiable factors. The direct entry models did not remove variables from the model based on their statistical significance. This was done to allow for a comparison of the models built using direct entry versus those built using backward elimination.

The backwards elimination models began with all the significant univariate factors in the model. Elimination of the candidate predictors in the backwards elimination models was determined by the Wald statistic. Candidate predictors with a p-value > 0.10 on the Wald test were removed from the model in a backwards stepwise fashion.

Multicollinearity was also assessed in the group of candidate predictors which demonstrated a significant univariate relationship with a successful outcome before being considered for inclusion in the final regression model. In instances where one candidate predictor is non-modifiable, than the non-modifiable predictor was included in the model and the modifiable predictor was excluded. If both candidate predictors that demonstrated high multicollinearity are modifiable, then the inclusion of one and the exclusion of the other was based on which variable has stronger clinically utility. If clinical utility was similar, than the candidate that has a greater body of research evidence on which to base clinical decision making was utilized.

In situations where a continuous predictor variable was found to be a significant univariate predictor of outcome, ROC curve analysis was used to dichotomize the continuous variable at a cut point of the continuous variable that maximized the Youden Index. The Youden Index "reflects the intention of maximizing overall correct classification rates and thus minimizing misclassification rates" when determining a cut point for a continuous predictor variable.[94] Two continuous predictor variables, the FABQ_work subscale and internal rotation strength as a percentage of the uninvolved side were significant univariate predictors of outcome. Therefore, they were dichotomized using ROC curve analysis and the Youden Index. The optimized cut-point for the FABQ_work subscale was found to be a score of 25. This cut-point resulted in a sensitivity of .93 and a specificity of .63 for correctly classifying those who reported a successful versus unsuccessful outcome, respectively. For internal rotation strength as a percentage of the uninvolved side, the optimal cutpoint was determined to be 76%. This cut-point resulted had a sensitivity of .61 and a specificity of .78. The direct entry model (with all 5 predictors) and the backwards elimination model were reevaluated using the dichotomous predictors, rather than the continuous versions of the FABQ_work score and internal rotation strength. These models were run to determine if dichotomizing the continuous predictors changed the diagnostic accuracy of the model compared to using the continuous analogue of the candidate predictor. By building a direct entry model that contained all the significant univariate predictors as well as a reduced model that contained only those predictors that had a Wald score with a p-value $< .10$ it was possible to compare the Nagelkerke R^2 values and discriminative abilities of the various models.

The individual probabilities derived from the logistic regression models were analyzed using ROC curves to calculate the area under the curve.[95] Larger values for the area under the curve indicate better discriminative accuracy of the model.[96]

3.2.4.2 Linear Regression Analysis

Due to the exploratory nature of this study, multiple linear regression was also used to determine the predictors of the change in the WORC score from baseline to 6-month follow-up. All the potential predictor variables that were analyzed using logistic regression were also analyzed for their univariate relationship to the WORC change score using linear regression. All potential predictors that had a significant relationship (p-value < .05) with the WORC change score on univariate regression were retained for building the final regression model. From this subset of predictors, non-modifiable predictors were added to the model sequentially based on the strength of their univariate relationship. Once non-modifiable factors no longer produced a statistically significant increase in the R^2 change statistic, the potentially modifiable factors were added to the model sequentially based on the strength of their univariate relationship to the WORC change score. Again, once additional potentially modifiable factors no longer produced a significant change in the R^2 change statistic, the model was considered complete. If a predictor in the model became non-significant when a potential predictor was added to the model, then the non-significant predictor was removed from the model and the effect its removal had on the R^2 value was calculated.

3.3 RESULTS

Complete baseline and 6-month follow up data were available for 46 of the 62 subjects (74% follow-up). 28 subjects met the criteria for being classified as a responder to surgery and 18 subjects failed to meet the criteria and were classified as nonresponders. A comparison of subjects who completed the study to those who dropped out is provided in Table 7. The mean age of individuals in the final sample was 46.4 ± 10.1 (years) and 63% were female. All subjects underwent arthroscopic subacromial decompression as well as the following operative procedures: 9 subjects (20%) had isolated subacromial decompression, 21 subjects (46%) had arthroscopic repair of a small cuff tear, 8 subjects (17%) had repair of a small rotator cuff tear as well as a labral repair or biceps tenodesis, 3 subjects (7%) had isolated labral repair, 2 subjects (4%) had a labral repair and biceps tenodesis, and 3 subjects (7%) underwent repair for small tears of both the supraspinatus and infraspinatus tendons. The severity of the rotator cuff tear (Table 8) was not a univariate predictor of PRO.

Of the predictor variables collected preoperatively, 3 of the factors from the patient characteristics and history of the shoulder condition (Table 9) were significant univariate predictors of PRO at 6 months. If a subject was on modified job duty (OR=.17, 95% CI: .03-.94), or had a worker's compensation claim (OR=.08, 95% CI: .01-.74) these factors decreased the probability of being a responder while surgery on the dominant shoulder (OR=11.96, 95%CI: 2.91-49.18) increased the probability of being a responder to surgical intervention.

From the baseline impairment measurements (Table 10), only impaired internal rotation strength was a significant predictor of PRO at 6 months. Individuals that had internal rotation strength of the involved side of at least 76% that of the uninvolved shoulder had an increased odds of being classified as a responder (OR=4.33, 95%CI: 1.21-15.4).

Among the patient-reported surveys (Table 11), which included psychosocial assessments, shoulder activity level and both the WORC and DASH, only the FABQ total score (OR=.95, 95%CI .91-.98) and the FABQ_work subscale (OR=.92, 95%CI .87-.97) were significant univariate predictors of outcome. These odds ratios are based on the continuous FABQ and FABQ_work subscale scores. ROC curve analysis determined a cut point of 25 on the FABQ_work subscale maximized the sensitivity and specificity of predicting a responder from a non-responder and the odds ratio for the dichotomized FABQ_work score was 22.3 (95% CI 4.0 – 123).

Table 12 presents the multicollinearity among the final group of potential predictors .A high degree of multicollinearity was found between the FABQ total score and the FABQ_work subscale ($r=.94, p<0.01$). Given the propensity of previous research to report the FABQ_work subscale as opposed to the total score and given the high degree of multicollinearity between the two measures, we choose to include the FABQ_work subscale in the logistic regression analysis and eliminated the FABQ total score from further consideration. There were also moderate levels of multicollinearity between the FABQ_work subscale score and the presence of modified job duty and a worker's compensation claim. Worker's compensation and modified job duty had a correlation of $r = 0.60$ indicating a moderate degree of correlation between these two variables as well.

3.3.1 Results of logistic regression model building using direct entry

The direct entry logistic regression model building process was described in the methods section (Section 3.2.4.1) and the results are presented in Table 13. In the direct entry model building process, non-modifiable factors such as surgery on the dominant shoulder, worker's compensation status and modified job duty status were added sequentially to determine the effect of each additional predictor on the model's discriminative ability and on the model's Nagelkerke R^2. Once the non-modifiable candidate predictors were added to the model, the two remaining candidate predictors based on univariate analysis, the FABQ_work subscale and IR strength as a percentage of the uninvolved side were added sequentially to the model. When all 5 of the predictor variables that had a significant univariate relationship were entered into the logistic regression model (Model 5 on Table 13), only dominant arm remained significant at the level of $p < .05$ (95% CI of the odds ratio: 3 to 153) while the FABQ_work subscale variable remained a significant predictor at the level p-value < .10 (95% CI of the odds ratio: .85-1.00).

Following the direct entry model building process, four additional models were built to determine whether a full model (Table 14 Model 1), with all the significant univariate predictors included, had better discriminative ability and a higher Nagelkerke R^2 than a reduced model that contained only predictors with a Wald test p-value <.10 (Table 14 Model 2). In addition to these two models, an additional full model using the dichotomized FABQ_work subscale variable and the dichotomized internal rotation strength variable was built (Table 14 Model 3) as well as a reduced model (Table 14 Model 4) that utilized the dichotomous predictors and contained only predictors with a Wald test p-value <.10.

The models in Table 14 indicate that the full model, utilizing continuous versions of the predictor variables (Model 1 Table 14) had the highest Nagelkerke R^2 ($R^2 = 0.66$) and the greatest discriminative ability based on an ROC curve area under the curve of 0.92. However, in this model, only the FABQ_work subscale and dominant shoulder variables remained significant predictors of outcome. The remaining 3 variables did not have a Wald test p-value < .10. When Model 2 was used to examine a reduced model with only the two significant predictors remaining, the Nagelkerke R^2 value was 0.61 and the area under the ROC curve was 0.87. Models 3 and 4 contain the dichotomized FABQ_work subscale and internal rotation strength predictors and the full model containing dichotomized analogs of the continuous variables (Model 3) had a Nagelkerke R^2 of 0.60 and an area under the ROC curve of 0.89. The reduced model (Model 4) with only the two variables significant at the Wald test p-value level < .10 had a Nagelkerke R^2 of 0.55 and an area under the ROC curve of 0.86.

All the models in Table 14 have overlapping 95% confidence intervals for the area under the ROC curve. However, both the full models contain 3 potential predictors that are not statistically significant predictors of outcome at a liberal p-value level of 0.10 on the Wald test. Model 4 contains the 2 significant dichotomized predictors and has an area under the ROC curve (AUC=.86) that is nearly identical to Model 2 (AUC = .87). Model 4 is identical to Model 2 except Model 2 contains the continuous analog of the FABQ_work subscale predictor.

Based on these results, Model 4 produces nearly identical discriminative ability to Model 2, and contains only predictors significant at a Wald test p-value level of < 0.10. The betas, odds ratios and their 95% confidence intervals for Model 4 are presented in Table 15. Based on Model 4, post-test probabilities of being a responder to surgery are dependent on whether an individual had surgery on the dominant arm and their FABQ_work subscale score being either less than or greater than 25. The model based post-test probabilities of being a responder, based on the two significant predictors (dominant arm and FABQ_work subscale score < 25) are presented in Table 16. The diagnostic accuracy statistics (sensitivity, specificity, likelihood ratios) of determining response to surgery based on the predictors in Model 4 (surgery on the dominant shoulder and FABQ_work subscale score < 25) are presented in Table 17.

3.3.2 Results for Linear Regression Analyses

The candidate predictors with a univariate regression p-value < .10 for predicting the WORC change score are listed in Table 18. Using the WORC change score as the dependent variable, linear regression analyses resulted in a final model (Model 5 Table 19) that included the following 3 predictors: 1) Dominant arm 2) Modified job duty 3) Age. The final model (Model 5 Table 19) with 3 predictors produced a R^2 of .50 and an adjusted R^2 of .46. Based on the model building strategy presented in the methods section, having surgery on the dominant shoulder was entered first into the regression model (Model 1 Table 19) since it had the strongest univariate relationship with the WORC change score. This resulted in an R^2 value of 0.28. Adding the baseline WORC score increased the R^2 value to .35. The baseline WORC score was a significant predictor of the WORC change score at the liberal p-value of .10 (but not at the .05 level). Next, the variable modified job duty was added to the model and resulted in an increase in the R^2 value of the model to .45 (Model 3 Table 19). Age was next added to the model, and it was an significant predictor of the WORC change score and the R^2 value increased from .45 to .52 (Model 4 Table 19). As age was added to the model, the WORC baseline score became a non-significant predictor at the p-value level of .10 (Model 4 Table 19). Therefore, the model was recalculated with age included but the WORC baseline score removed (Model 5 Table 19). The removal of the WORC baseline score resulted in a R^2 decrease of only .02, which was not a statistically significant decrease. Therefore, the WORC baseline score could be removed from the model without any deleterious effects to the models predictive ability. The remaining univariate predictors were entered sequentially into Model 5, but none improved the R^2 value of the model or were significant predictors of the WORC change score. Therefore, Model 5 (Table 19) was the most parsimonious model available.

Evaluation of the assumptions for the model revealed that there was: 1) no significant departure from a linear relationship between the individual predictor variables, the composite predictors, and the dependent variable (Figures 2-3); 2) homoscedasticity appears to be met as the residuals are distributed fairly evenly across the predicted values of the dependent variable

(Figure 2); 3) Multicollinearity was not found in the final 3 predictor variables and the variance inflation factors (VIFs) were between 1.03 and 1.14 indicating no negative effects of multicollinearity on the final model; 4) Outliers, high leverage points and highly influential points were examined. 2 cases had a standardized residual of 2.6 from their predicted WORC change scores. However, data was correctly entered for these subjects and there is no reason to exclude their data from analysis. All leverage points were less than 0.30 and all Cook's distance values were less than .12 indicating no exaggerated influence of one particular subject's data. 5) Normality of the residuals was examined visually using the P-P plot (Figure 4) The plot does not indicate a significant departure from a normal distribution of the residuals.

3.4 DISCUSSION

Utilizing information readily collected during the examination of an individual who requires surgery for rotator cuff pathology, it was possible to accurately predict whether or not that individual will report a successful outcome 6 months following surgery. Both direct entry and backwards elimination logistic regression models were evaluated and in all the models (Tables 13 -14), surgery on the dominant shoulder and having a low FABQ_work subscale score were the only predictors to remain statistically significant predictors of successful outcome. In Table 13, the 1st model consisted of only surgery on the dominant shoulder as a predictor and it produced a model with an area under the ROC curve of 0.77 indicating good discrimination between those who were classified as a success versus those who were not. While adding the additional predictor variables did result in small increases in the discriminative ability of the model, all but the FABQ_work subscale score failed to remain statistically significant predictors of outcome.

Since the FABQ_work subscale is a continuous measure, dichotomizing it allows for easier use of the final model by clinicians by allowing them to score a patient as above or below the cut-score on the FABQ_work subscale and as having surgery on their dominant or non-dominant shoulder. Whether or not dichotomizing the FABQ_work subscale negatively impacted the accuracy of the model was evaluated (Table 14). In Table 14, Model 2 and Model 4 began with the same 5 candidate predictors with the exception of Model 2 using the continuous FABQ_work subscale and Model 4 using the dichotomized FABQ_work subscale. The discriminative ability of model 2 had an area under the ROC curve of .87, while Model 4 had an area under the ROC curve of .86. This minimal difference between the discriminative ability of the model with the continuous FABQ_work subscale predictor and the model with the dichotomized version of this variable indicates the dichotomized FABQ_work subscale variable can be used in the model without relevant loss of discriminative ability. Since it is simpler for clinicians to utilize the results of this model when the predictor variables are dichotomized, the final model did incorporate the dichotomized version of the FABQ_work subscale variable. The following paragraph provides an example of utilizing the likelihood ratios that can be calculated from the final model.

Given that there are only 4 possible combinations of the final 2 predictor variables, it is straightforward to calculate the likelihood ratios for these 4 combinations (Table 17). For instance, knowing that the patient is going to have surgery on their dominant shoulder and that their FABQ_work subscale score is less than 25 results in a post-test probability of successful outcome of 88% (Table 17). If this model is validated in a larger group of similar patients, then it could be used for shared decision making between the care providers and the patient to set expectations for their postoperative outcome. The likelihood ratios do indicate that the combination of surgery on the dominant shoulder and a FABQ_work score below 25, when both are found together, results in a positive likelihood ratio of 4.4 which produces a moderate shift of the pre-test probability for success from 61% to 88%. Not having both of these factors also produces a moderate shift towards not having a successful outcome based on the calculated negative likelihood ratio of 0.23. The very low positive likelihood ratio of .07, when surgery is on the non-dominant shoulder and the FABQ_work subscale score is greater than 25 produces a substantial drop in the probability of success from a pre-test probability of 61% to a post-test probability of only 10%. The remaining two combinations do little to alter the pre-test to post-test probability of a successful outcome. The lack of a large number of subjects within each possible combination of the predictor variables results in a model that may be difficult to reproduce due to poor model stability.

In a prior study, Gulotta et al[97] were unable to find pre-surgical predictors of successful PRO following arthroscopic rotator cuff repair. In their study, an American Shoulder and Elbow Surgeons score of > 90 points at the 5-year follow-up was used to determine a successful PRO. However, they did not measure psychosocial variables such as fear-avoidance. Fear-avoidance, in the logistic regression model, was found to be a predictor of outcome. Also, strength measurements of shoulder flexion and external rotation were quantified using the Medical Research Council's 6-point scale and they did not quantify internal rotation strength. Whether or not they examined the prognostic ability of surgery on the dominant versus non-dominant shoulder is unclear. In addition to illuminating the difficulty in predicting PROs, their work and the current study also provide an example of the varying criteria used to dichotomize individuals as a responder or a nonresponder For instance, the current study used the global rating of change score as well as the WORC change score to dichotomize the subjects into those who responded versus those that did not, while the study by Gulotta et al used a score of > 90 on the American Shoulder and Elbow Surgeons Score as their criteria by which to dichotomize their subjects and analyze predictors of PRO. These variations in determining outcome make it difficult to synthesize the current body of literature that has examined predictors of PROs.

3.4.1 Fear-avoidance as a predictor of response to rotator cuff surgery

Fear-avoidance was the only potentially modifiable factor of the candidate predictors to remain an significant predictor of PRO. Fear-avoidance has established itself as a predictor of outcome in patients with acute low back pain as well as in patients with chronic back pain who have undergone disc prosthesis surgery for degenerative disc disease.[76, 98, 99] The origins of the fear-avoidance model stem from the work of Lethem et al[100] and Slade et al.[101] They postulated that pain perception involves 2 components: a sensory-discriminative component and a motivational-affective component. The sensory-discriminative component is the physiological nociceptive input that occurs with tissue damage while the motivational-affective is the psychological component of pain perception. In individuals with what they termed "exaggerated pain perception," they proposed that there was a disconnect that develops between the two components of pain perception. For example, decreased nociceptive input should result in less pain sensation and thus an improved emotional response (i.e. pain behavior) to pain. In this example, the two components of pain perception move together in the same direction. In those who have an exaggerated pain response, these two components become dissociated. When dissociation of the sensory-discriminative and motivational-affective components occurs, an individual can have improving nociceptive input (sensory-discriminative component of pain) but still have worsening psychological effects (motivational-affective component). In their model of fear-avoidance, the psychological component is affected by factors such as stressful life events, personality, personal pain history, pain coping strategies, and fear of pain. Of these factors, they believe fear of pain to be the most important. As with any fear, an individual decides where on the continuum they fall in terms of their confrontation-avoidance of their fear of pain. Lethem writes, “A central tenet of the model is that ‘fear of pain’ generates in some individuals a strategy of avoidance rather than confrontation which in turns leads to both physical and psychological reinforcement of the invalid status.” In individuals requiring shoulder surgery, who have elevated fear-avoidance behaviors as quantified by the FABQ, it is possible that their elevated fear-avoidance behaviors contribute to poor patient-reported outcomes. According to Lethem and Slade’s model, these individuals with elevated fear-avoidance will fail to confront activities which may cause pain, and withdraw from work and activities of daily living. This in turn can worsen the motivational-affective component of pain perception. Then, even when the pathoanatomical structure is restored after surgery, there is now dissociation between the improving sensory-discriminative component of pain perception and the stagnant or worsening motivational-affective component of pain perception. In these individuals, they fail to report improved abilities and overall health following surgery because they still have a high level of pain perception stemming from the motivational-affective component of pain perception.

While we are not aware of prior studies examining the role of fear-avoidance in individuals undergoing rotator cuff surgery, there have been studies that have examined fear-avoidance in patients receiving treatment for shoulder pain.[102-104] In these studies the predictive ability of fear-avoidance has been equivocal with some studies reporting fear-avoidance as a predictor of outcome[103] and others reporting it was not.[102, 104] However, these studies lack homogeneity in the instrument used to quantify fear-avoidance and in the criterion used to measure outcome. Therefore, comparison of the prognostic ability of fear-avoidance is limited until future studies are performed that utilize the same instruments to measure fear-avoidance and to measure patient-reported outcome. The current study provides an initial study indicating that the FABQ may have prognostic accuracy in a population similar to the one presently examined when the same criteria for success are utilized.

Although currently unexplored in individuals with rotator cuff pathology, there exists evidence in the treatment of patients with low back pain, that fear-avoidance-based physical therapy results in superior outcomes when applied to individuals with acute low back pain who also have high fear-avoidance scores at baseline.[105] Fear-avoidance based treatments for low back pain include elements of patient education that downplay the role of anatomical findings and encourages the patient to confront the activities which cause them fear.[106, 107] The prescription of therapeutic exercise and activities are guided using predetermined quotas in the activity or the intensity and repetition of exercise. Positive reinforcement is provided when the quota is met and a new quota is set. Similar to the work done in low back pain, development and evaluation of fear-avoidance based treatment protocols should be considered as a possible strategy for improving outcomes in individuals with high FABQ scores who undergo rotator cuff surgery if elevated FABQ scores are consistently found to predict a less favorable outcome.

Fear-avoidance may be a moderator of outcome in individuals having rotator cuff surgery. Being a moderator indicates that fear-avoidance may be a mechanism that can cause or prevent a successful outcome and by intervening to improve fear-avoidance, treatment outcomes may be improved.[108] To examine whether or not fear-avoidance has a moderating effect on the outcome of individuals having rotator cuff surgery, future randomized trials would need to be structured to detect the presence of fear-avoidance as a moderating factor and then treat the fear-avoidance in one treatment group while not addressing it in the control group.[108] To do this, individuals undergoing rotator cuff surgery would be measured on their level of fear-avoidance preoperatively and then randomized to either a standard of care preoperative/postoperative program or a program aimed to mitigate the suspected negative effect of high fear-avoidance levels on outcome. The hypothesis would be that individuals with high fear-avoidance levels preoperatively would benefit more by being in the treatment group focused on addressing their fear-avoidance than would someone without high fear-avoidance levels. The benefit would be measureable in the outcome of surgery as well as in the improvement of their fear-avoidance levels. In contrast, those with high levels of fear-avoidance who were randomized to the standard of care group, which would do nothing to address their elevated fear-avoidance, would have no improvement in their fear-avoidance levels and a worse outcome with the surgery. In this scenario, fear-avoidance levels may be shown to be a moderator of rotator cuff surgery.

3.4.2 Surgery on the dominant arm as a predictor of response

The fact that surgical intervention on the dominant shoulder was so strongly predictive of a positive response to surgery was somewhat unexpected. Recent studies that have examined whether or not rotator cuff repair on the dominant arm predicts clinical outcome have reported that it does not.[4, 14, 15, 38] Patel et al also reported that individuals undergoing arthroscopic subacromial decompression on their non-dominant shoulder had greater improvements on their Constant score than did those who had surgery on the dominant shoulder.[17] They proposed that surgery on the non-dominant shoulder would most likely lead to greater improvements in self-reported function because less demands are placed upon the non-dominant shoulder as compared to the dominant one. However, upon closer examination of studies that have shown no relationship between surgery on the dominant shoulder and outcome, it is evident that the outcomes were mainly impairment-based (such as the Constant score and UCLA score) and none of the studies used the WORC or DASH as a patient-reported functional outcome.[4, 14, 15, 38] While items measured on the Constant score and UCLA shoulder score such as shoulder strength and ROM may not differ between dominant and non-dominant shoulders that undergo cuff surgery, a patient-reported outcome would seem more likely to demonstrate larger improvements on a dominant shoulder that undergoes surgery as compared to a non-dominant shoulder. This may be due to the fact that when the dominant arm is affected, functional limitations become much more evident to the patient than when the non-dominant arm is involved. Therefore, surgery on the dominant shoulder would logically result in larger improvements in functional activities of the upper extremity. In fact, those subjects who had surgery on their dominant arm started with a mean WORC score of approximately 41, while those subjects having surgery on the non-dominant arm had a baseline WORC score of 48. Therefore, having the shoulder pathology on the dominant side resulted in more self-reported limitations at baseline.

3.4.3 Impairments of the shoulder complex as potential predictors

Due to a lack of the studies examining the prognostic ability of shoulder motion, strength and scapular dyskinesis in this population, we examined several of these measures for their prognostic ability. Loss of glenohumeral internal rotation has been reported in patients with subacromial impingement syndrome and can be caused by tightness in the posterior capsule.[61-63] Tightness of the posterior capsule, in turn, has been shown to increase anterior and superior displacement of the humeral head on the glenoid fossa.[61, 109-111] The increased anterior and superior displacement of the humeral head secondary to posterior capsular tightness is believed to result in impingement symptoms and, eventually, rotator cuff tears.[112] While these biomechanical links of posterior capsular tightness, subsequent loss of internal rotation motion and altered humeral head position have been demonstrated, our results were unable to demonstrate that loss of glenohumeral internal rotation range of motion was a predictor of outcome following rotator cuff surgery.

While loss of glenohumeral external rotator strength has been demonstrated in individuals with subacromial impingement[64, 65] , we were unable to find any recent studies that examined the ability of rotator cuff strength to predict PRO after rotator cuff surgery. We collected measures of internal and external rotation strength as well as elevation in the scapular plane using a handheld dynamometer to maximize clinical utility of the measurements. However, none of the strength measurements were found to be independent predictors of outcome.

As with shoulder motion and strength, there is a paucity of studies that examine the prognostic ability of scapular dyskinesis in individuals similar to those in the current study. Although we dichotomized scapular motion in these individuals as normal versus dysfunctional (medial or inferior border winging, scapular elevation), our results did not indicate that the presence or absence of scapular dyskinesis was a prognostic factor for outcome 6 months after arthroscopic surgery for impingent with or without a rotator cuff repair.

3.4.4 Linear regression for predicting the WORC change score

Additional linear regression analyses were carried out to determine which candidate predictors were significant predictors of the WORC change score. Due to the exploratory nature of this study and the limited sample size, the linear regression results allowed for a comparison to the results that were found with the logistic regression analyses. The change in the WORC score from baseline to 6-month follow up was selected as the dependent variable for the linear regression, rather than the WORC score at 6-months, because the criteria for being a responder in the logistic regression analyses was highly dependent on having a specific change score on the WORC. Therefore, the WORC change score was more closely related to the criteria used for outcome in the logistic regression analyses than was the WORC score at 6-months. From the linear regression analyses, it was determined that surgery on the dominant arm, being on modified job duty, and the subject's age at baseline were the 3 independent predictors of the WORC change score. The strongest predictor of the WORC change score was having surgery on the dominant arm (standardized beta of .46) followed by age (standardized beta of -.42) and then being on modified job duty (standardized beta of -.37). The adjusted R-squared for the final model was .46 (Model 5 table 19). The fact that the model only had an adjusted R^2 of .46 indicates that a majority of the variance in the WORC change score was not explained by the predictors in that model. A full model, containing all seven predictors that had a significant univariate relationship to the WORC change score, was constructed to determine how much variance was explained by the full model. The full model (Full Model Table 19) had an adjusted R^2 of only .44, indicating slightly worse predictive ability in the full model than in the final model (Model 5 Table 19).

Surgery on the dominant arm was a significant predictor in both the logistic and the linear models. Subjects who had surgery on the dominant arm, versus the non-dominant arm, tended to have greater improvements in their WORC score. Surgery on the dominant arm resulted in an average WORC score increase of approximately 40 points as opposed to only 17 points for surgery on the non-dominant arm. As discussed in section 3.4.2, surgery on the dominant arm, as a predictor of outcome has been explored mainly using outcome measures such as the Constant score and the UCLA score which are heavily weighted to measure impairments of shoulder strength and motion. The WORC is a self-reported functional outcome scale and there are no prior studies that have investigated predictors of its change in this population. From our results, it appears that the WORC change can be expected to be greater when surgery is on the dominant arm.

The subject being on modified job duty was a predictor in the final linear regression model. We are unaware of any previous studies that have examined this variable as a predictor of outcome in this population. While it was also a significant univariate predictor of response in the logistic model, it was not a final predictor in the logistic model. The moderate level of collinearity that exists among the variables: modified job duty; worker's compensation; and FABQ_work subscale indicates that these variables share variance (Table 12). Also, of the 7 subjects who had a worker's compensation claim, 5 of these were subjects on modified job duty (out of only 7 total subjects on modified job duty). Therefore, there is significant overlap in the individuals who comprise the predictors of modified job duty and worker's compensation. Once one of these two variables was in the model, the other was superfluous as it explained somewhat overlapping variance. The FABQ_work subscale exhibited a similar situation in that it was also moderately correlated to the predictors worker's compensation and modified job duty. While

FABQ_work subscale was a final predictor in the logistic model, it was only a univariate predictor in the linear regression model. Since modified job duty was entered into the linear model prior to the FABQ_work score, and the variables shared variance, the FABQ_work score was not a significant predictor of the WORC change score while modified job duty was.

Modified job duty was a predictor of the WORC change score, but not of being a responder in the logistic model. The opposite was true of the FABQ_work score. It remained a significant predictor in the logistic model but not in the linear model. Therefore, the choice of outcome, although similar in the two models, results in correlated, but different predictors remaining in the final models. As was mentioned, the subjects who were on modified job duty were predominantly the same individuals who had a worker's compensation claim. Therefore, in the linear model, modified job duty is serving as somewhat of a proxy for having a worker's compensation claim. Having a worker's compensation claim, was shown in our systematic review, to be a predictor of less successful outcomes following rotator cuff surgery when the DASH was used the outcome measure. To date, there are no previous studies that have examined worker's compensation as a predictor of the WORC or change in WORC in this population.

Age, was an independent predictor of the WORC change score. Increasing age predicted less improvement on the WORC. In contrast to this result, age was not a predictor of response in the logistic model. As was determined by our systematic review, age, as a predictor of outcome, has been equivocal to this point. Again, out study shows that age as a predictor is dependent on the outcome used to measure change or response to surgery. The logistic model required a 17 point improvement on the WORC as part of its criteria for being classified as a responder. However, the second part of the criteria was that the subject needed to report they were at least "quite a bit better" at the 6-month time point. Therefore, increasing age resulted in less improvement on the WORC score, but not in terms of satisfaction with the outcome of surgery as measured by the global rating of change score. It appears that in spite of less improvement on the WORC, older subjects still reported that they were "quite a bit better". A similar finding was reported by Razmjou et al.[113] They argued that lower expectations and lower physical demands of the shoulder in older individuals who have cuff surgery may result in greater satisfaction in older individuals.

3.5 LIMITATIONS

The small sample size used in this study results in several important limitations. First, the complexity of shoulder injuries, often involving several structures, necessitates future studies with large samples to examine the many pathologies that can occur simultaneously in the injured shoulder and potentially impact self-reported outcomes.

Second, a lack of statistical power and collinearity among predictors are strong potential explanations for why worker's compensation status, and other potential predictors found to be non-significant, were not included in the final logistic regression model or the final linear regression model. For instance, only 7 of the 46 subjects had worker's compensation claims. The low percentage of subjects with a worker's compensation claim exacerbates the problem of lack of power. In fact, the descriptive data shows that only 1 of the 7 subjects with a worker's compensation claim had a successful outcome. If a larger sample of subjects was used, the fact that only 14% (1 out of 7) of those with a worker's compensation claim compared to 69% (27 out of 39) of those without a worker's compensation claim had a successful outcome, having a worker's compensation claim would have most likely remained a significant predictor of outcome. In addition, the moderate to moderately-strong correlations among the FABQ_work subscale score, modified job duty status, and worker's compensation status indicated that these potential predictors were competing with one another to enter the regression models. Again, the descriptive data shows that of the 7 subjects having a worker's compensation claim, 5 of them were also the subjects with modified job duty (and only 2 other subjects were on modified job duty).

Third, as with any prognostic study, the lack of a gold standard by which to determine a "successful" outcome makes the results of our study specific to the criteria we utilized.. The fact that we required our subjects to report a 17-point improvement in their WORC score as well as a +5 or better (at least "quite a bit better") on their global rating of change to be classified as a responder undoubtedly lowered our rate of "responders" as compared to what is often published in the literature. However, these strict criteria help to insure that if this model, once validated, predicts a "successful" outcome; you can have a high level of confidence that the result of surgery will be favorable. This study included a small number of subjects, explored a large number of candidate predictors, and should be viewed as an initial derivation of a predictive model. Small datasets are prone to overfitting and the results found in this cohort need to be replicated in a larger prospective cohort study prior to attempting to make clinical decisions based on this model.[114, 115]

Finally, there were differences in baseline measures between those individuals who completed the study and those who dropped out of the study. The CES_D, used to quantify depressive symptoms, demonstrated a statistically significant difference between those that completed the study and those that did not. The group who completed the study had a lower mean CES_D score (mean = 6) while those not completing the study had a CES_D mean score of 12. The averages indicate that those who dropped out of the study had higher depressive symptomatology than those who completed the study. Since depression has been linked to worse self-reported outcomes[116] and thus more failures when self-reported measures are used as the criterion for success, it is plausible that the loss to follow-up of those with higher levels of depressive symptoms increased the percentage of individuals achieving a successful outcome. In addition, the loss of those with higher CES_D scores may have caused this potential predictor to be mistakenly eliminated as a predictor of outcome since those who dropped out could not be analyzed in the final model.

3.6 CONCLUSIONS

Although there are limitations, the current findings regarding the predictive ability of the FABQ_work subscale in the logistic model is intriguing and speaks to the potential impact of psychosocial variables on patient-reported outcome after arthroscopic subacromial decompression without or with rotator cuff repair. When change in the WORC score is used as the outcome or is integral in the criteria for success, as it was in the logistic modeling part of this study, it appears that having surgery on the dominant shoulder predicts favorable patient-reported outcomes and larger improvements in the WORC score at 6-months following surgery. With modified job duty being predictive in the linear model and high FABQ_work scores being predictive in the logistic model, both models found work related factors to be predictive of lesser outcome which is consistent with prior research. It also demonstrates that predictors can vary based on the specific outcome used to measure change or success. Finally, age did predict less improvement on the WORC, but it did not predict whether a subject met the criteria for success in the logistic model. Therefore, older subjects, despite experiencing less change on the WORC still tended to report that they were at least “quite a bit better” on the global rating of change at the 6 month time point. Caution in interpreting these results is needed until further studies are done in larger, similar samples using the same outcome measures to test the robustness of the predictors.

3.7 TABLES

Table 7 Comparison of characteristics of individuals who completed the study versus dropouts

	Complete Data (n=46)	Dropout (n=16)	p-value of Mann-Whitney U test
Age (mean)	45	46	.66
WORC baseline	43	34	.08
DASH baseline	40	47	.12
FABQ_work subscale	17	23	.14
CES_D	6	12	.04*
Beck	7	8	.79
IR strength as % NI	73	73	.95
Supraspinatus strength as a % NI	67	57	.35
Infraspinatus strength as a % of the NI	74	66	.26

*- p-value < 0.05; abbreviations: NI, non-involved side

Table 8 Univariate relationship of intraoperative findings of rotator cuff condition to outcome

		Responder (n=28)	Nonresponder (n=18)	Odds Ratio	OR 95% Confidence Interval	OR p-value
Intraoperative Cuff Status	Severity of Rotator Cuff Injury					
	No tear, n (%)	9 (32%)	3 (17%)	Ref	Ref	.52
	Partial tear	15 (54%)	12 (67%)	.42	.09-1.89	.26
	Full tear	4 (14%)	3 (17%)	.44	.06-3.24	.42

Abbreviations: Ref, reference group

Table 9 Univariate relationship of patient characteristics and shoulder history to outcome

		Responders (n=28)	Nonresponders (n=18)	Odds Ratio (OR)	OR 95% Confidence Interval	OR p-value
Patient Characteristics	Age (mean ±SD)	46 ±8	47 ±13	.98	.93-1.04	.61
	Female, n (%within group)	18 (64%)	11 (61%)	1.31	.40-4.32	.66
	Level of Education					
	High school	2 (7%)	1 (6%)	Ref	Ref	.97
	Some college	9 (32%)	7 (39%)	.64	.05-8.62	.74
	College graduate	8 (29%)	6 (33%)	.67	.05-9.19	.76
	Post-graduate work	7 (25%)	4 (22%)	.92	.06-12.98	.92
	Employment Status					
	Modified Job Duty	2 (7%)	5 (28%)	.17	.03-.94	.04*
	WC case	1 (4%)	6 (33%)	.08	.01-.74	.03*
	Current Smoker	6 (21%)	5 (28%)	.78	.20-3.09	.72
Characteristics of Shoulder Condition	Duration of Symptoms					
	3 to 6 months	6 (21%)	8 (44%)	Ref	Ref	.19
	7 to 12 months	7 (25%)	2 (11%)	4.67	.70-31.04	.11
	13 to 24 months	7 (25%)	7 (39%)	1.33	.30-5.91	.71
	25 months	6 (21%)	1 (6%)	8.00	.75-85.31	.09
	Dominant Shoulder	23 (82%)	5 (28%)	11.96	2.91-49.18	.001*
	Known Cause of Pain	15 (54%)	14 (78%)	.39	.10-1.51	.17
	Able to Sleep on Shoulder	6 (21%)	3 (17%)	1.5	.32-6.99	.61
	Pain past the elbow	12 (43%)	5 (28%)	2.23	.62-8.08	.22
Trt	PT Treatment	19 (68%)	13 (72%)	1.04	.27-4.02	.95
	Cortisone Injection	15 (54%)	14 (78%)	.39	.10-1.51	.17

* - p=value < .05

Abbreviations: WC, Worker's Compensation; PT, physical therapy treatments; OR, odds ratio; Trt, Treatment; Ref, reference group

Table 10 Univariate relationship of impairment measures to outcome

			Responder (n=28)	Nonresponder (n=18)	Odds Ratio	OR 95% CI	OR p-value
Impairment	Pain	NPRS Pain Levels					
		Current	5.1 ±2.4	5.5 ±1.8	.91	.68-1.21	.53
		Worst	7.2 ±1.9	7.2 ±2.2	1.02	.75-1.38	.92
		Best	3.1 ±2.3	3.3 ±2.3	.96	.73-1.25	.75
	Scapula	+ SAT	5 (18%)	4 (22%)	.76	.17-3.32	.72
		+Scapular Dyskinesis	13 (46%)	8 (44%)	1.08	.33-3.56	.90
		Scapular Index Score	75.5 ± 7.6	77.6 ±9.9	.97	.91-1.04	.43
	AROM	Shoulder Flexion AROM(°)	146° ±22	137° ±33	1.01	.99-1.04	.31
		Glenohumeral IR AROM (°)	66° ± 18	64° ±21	1.00	.97-1.04	.78
	Strength	Supraspinatus strength as a % of the uninvolved side	67 ±25	67 ±31	.94	.11-8.35	.96
		Infraspinatus strength as a % of the uninvolved side	77 ±23	69 ±22	4.76	.30-74.42	.27
		Internal rotation strength as a % of NI	80 ±24	64 ±17	38.48	1.58-940	.02*
		IR strength ≥ 76% of NI	17 (61%)	5 (28%)	4.33	1.21-15.4	.02*
		ER/IR Strength Ratio	.73 ±.26	.84 ±.30	.23	.02-2.20	.20

*- p-value<.05

Abbreviations: NPRS, numerical pain rating scale; SAT, Scapular Assist Test; AROM, active range of motion; IR, internal rotation; ER, external rotation; NI, non-involved side

Table 11 Univariate relationships of patient-reported instruments to outcome

		Responder n=28	Nonresponder n=18	Odds Ratio	OR 95% CI	OR p-value
Psychosocial Surveys	Beck Anxiety Index	6 ±8	9 ±9	.96	.90-1.04	.32
	FABQ					
	Total Score	35 ±20	60 ±21	.95	.91-.98	.002*
	Work Subscale	11 ±11	26 ±15	.92	.87-.97	.001*
	Work subscale ≤ 25	26 (93%)	7 (39%)	22.29	4.02-123	.000*
	Physical activity scale	17 ±5	19 ±6	.93	.83-1.04	.19
	CES-D	6 ±5	8 ±9	.96	.88-1.04	.34
WORC DASH SAS	Baseline WORC Score	45 ±18	41 ±22	1.01	.98-1.04	.55
	Baseline DASH Score	38 ±15	44 ±20	.98	.94-1.02	.27
	Shoulder Activity Scale	13 ±5	13 ±4	1.01	.89-1.16	.85

*- p-value<.05

Abbreviations: FABQ, Fear Avoidance beliefs Questionnaire; CES-D, Center for Epidemiologic Studies Depression Scale; WORC, Western Ontario Rotator Cuff Index; DASH, Disabilities of the Arm, Shoulder, and Hand Outcome Measure; SAS, Shoulder Activity Scale

Table 12 Multicollinearity among candidate predictors

	FABQ	FABQ-work	FABQ-work < 25	IR strength as a % of NI side	IR strength ≥ 76% of NI	Modified Job Duty	WC Case	Dominant arm involved
FABQ	1	.94	-.71	-.29	-.34	.45	.61	-.29
FABQ_work		1	-.78	-.31	-.39	.49	.58	-.31
FABQ_work < 25			1	.25	.23	.48	-.68	-.48
IR strength as a % of NI side				1	.86	-.23	.30	.13
IR strength ≥ 76% of NI					1	-.11	-.17	.14
Modified Job Duty						1	.60	-.10
WC Case							1	-.28
Dominant Arm Involved								1

Abbreviations:IR, internal rotation; WC, worker's compensation case; NI, non-involved side

Table 13 Logistic Regression Model Building Process

<u>Criteria for Success:</u> 1. Global rating of change ≥ +5 2. At least a 17 point improvement on the WORC from baseline to 6-months postop.					
Variables entered based on univariate analysis. <u>Non-modifiable variables were entered first</u> Order of Entry into Model: 1. surgery on dominant shoulder 2. work comp status 3. modified job duty 4. baseline fabq_work 5. IR strength as a % of non-involved side					
	Odds ratio	95% CI for Odds Ratio	Nagelkerke R^2	Discriminative ability of the model (Area under ROC curve)	95% CI for the AUC
Model 1: Dominant Arm	11.96	3-49	.36	.77	.62-.92
Model 2: Dominant Arm Worker's Compensation	10.4 .10	2 – 46 .008 – 1.14	.45	.81	.68-.95
Model 3: Dominant Arm Worker's Compensation Modified Job Duty	12.6 .29 .20	3 – 63 .02 – 4.6 .02 – 2.3	.48	.84	.72-.96
Model 4:Dominant Arm Worker's Compensation Modified Job Duty FABQ_work subscale	11.3 .81 .37 .94	2 – 59 .04 – 16.7 .03 – 3.9 .87 – 1.01	.54	.88	.77-.99
Model 5:Dominant Arm Worker's Compensation Modified Job Duty FABQ_work subscale IR strength as a % of the non-involved side	20.4 1.29 .30 .92 16.7	3 – 153 .06 – 30 .03 – 3.6 .85 – 1.00 .14 - 2023	.66	.92	.84-1.00

Abbreviations: IR, internal rotation of the glenohumeral joint; AUC, area under the ROC curve

Table 14 Logistic Models: Non-modified Predictors versus Dichotomized Predictors

	Model 1 and 2: non-modified predictors		Model 3 and 4: continuous predictors dichotomized	
Criteria for success	1. Global change ≥ +5 2. 17 point change on WORC		1. Global change ≥ +5 2. 17 point change on WORC	
Variables entered based on univariate analysis	1. baseline FABQ_work subscale 2. workers compensation status 3. modified job duty 4. IR strength as a % of non-involved 5. surgery on dominant shoulder		Same as Model 1, except FABQ_work subscale and IR strength were dichotomized into FABQ_work scores >/<25 and IR strength >/< 76% of the uninvolved side.	
	Model 1	**Model 2**	**Model 3**	**Model 4**
Method of candidate predictor selection	Enter all 5 variables	Backward elimination using Wald value > .10 to exclude predictors	Enter all 5 variables	Backward elimination using Wald value > .10 to exclude predictors
Predictors with Wald test p-value <0.10	FABQ_work, Dominant shld	FABQ_work, Dominant shld	Dichotomized FABQ_work, Dominant shld	Dichotomized FABQ_work, Dominant shld
Nagelkerke R^2	.66	.61	.60	.55
Discriminative Ability of the Model (Area Under the Curve)	.92	.87	.89	.86
95% CI for the AUC	.83 – 1.00	.76 - .99	.78 – 1.00	.73 – 1.00

Abbreviations: IR, internal rotation of the glenohumeral joint; AUC, area under the ROC curve

Table 15 Final logistic regression model for predicting responders

Predictor Variable	Beta	OR	95% CI of OR	p-value
FABQ_work subscale ≤ 25	2.73	15.29	2.30-101.90	.005
Involvement of dominant arm	2.19	8.93	1.75-45.68	.009
Constant	-2.746	.064		

Abbreviations: FABQ, fear avoidance beliefs questionnaire; OR, odds ratio; CI, confidence interval

Table 16 Predicted probabilities of success based on final logistic regression model

Combinations of Predictor Variables	Formula to calculate Odds (from final model Table 15)	Odds	Probability (Probability = odds/(1+odds)
1. Dominant shoulder 2. FABQ_work < 25	e^2.73 + 2.19 – 2.746	8.8	8.8/9.8 = 90%
1. Dominant shoulder 2. FABQ_work > 25	e^0 + 2.19 -2.746	.57	.57/1.57 = 36%
1. Non-dominant shoulder 2. FABQ_work < 25	e^ 2.73 + 0 – 2.746	.98	.98/1.98 = 50%
1. Non-dominant shoulder 2. FABQ_work > 25	e^0 + 0 – 2.746	.06	.06/1.06 = 6%

Table 17 Sensitivity, Specificity, and Likelihood Ratios for the Predictor Variable Combinations

Predictor Combinations	Sensitivity	Specificity	Positive Likelihood Ratio	Post-test Probability of Success*	Negative Likelihood Ratio
Dominant Arm and FABQ_work ≤ 25	.75	.83	4.4	88%	0.23
Dominant Arm and FABQ_work > 25	.07	.89	.64	50%	1.04
Non-dominant Arm FABQ_work ≤ 25	.18	.78	.80	55%	1.06
Non-dominant Arm FABQ_work > 25	.03	.50	.07	10%	1.93

Legend: * - the pretest probability of success is the prevalence of success, 61%

Table 18 Candidate predictors of WORC change score

Candidate Predictor	R^2	p-value of univariate analysis
Non-modifiable candidate predictors		
Dominant arm	.28	.00
WORC score at baseline	.14	.01
Modified job duty	.11	.03
Age	.10	.05
Specific Cause	.08	.08
Potentially modifiable candidate predictors		
Worst pain (past 24 hours)	.17	.01
FABQ_work subscale	.08	.08

Abbreviations: NI, non-involved side, ER, external rotation; IR internal rotation

Table 19 Model Building Results for Prediction of WORC change score

Linear Regression Model Building

1. Enter non-modifiable candidate predictors from Table 18 into the model based on strength of the univariate analysis' R^2 value
2. Enter modifiable predictors from Table 18 into the model based on strength of the univariate analysis' R^2 value

- **FABQ and FABQ_work are highly correlated (r >.90) therefore FABQ_work was assessed for entry rather than FABQ to remain consistent with previous logistic regression model.**

	β (95%CI)	Standardized β	P for beta	R^2	Adj R^2	R^2 Change	P for change
Model 1: 1.Dominant arm 2. Constant	24.6 (12.2 to 37.0) 16.9	.53	<.001	.28	.26	na	<.001
Model 2: 1.Dominant arm 2.WORC baseline 3. Constant	21.6 (9.2 to 33.9) -.0.31 (-.62 to .002) 32.4	.46 -.26	.001 .051	.35	.32	.07	.051
Model 3: 1. Dominant arm 2. WORC baseline 3. Modified job duty 4. Constant	17.6 (5.7 to 29.4) -0.42 (-.72 to -.12) -20.7 (-36.1 to -5.3) 43.3	.38 -.36 -.34	.005 .007 .010	.45	.41	.10	.01
Model 4: 1. Dominant arm 2. WORC baseline 3. Modified job duty 4. Age 5. Constant	19.2 (7.8 to 30.5) -0.22 (-.55 to .12) -23.7 (-38.6 to -8.9) -0.74 (-1.39 to -.09) 68.5	.41 -.19 -.39 -.32	.002 .194 .003 .028	.52	.47	.05	.028
In Model 4, WORC baseline became a non-significant predictor of the WORC change score. Therefore, it was removed from the Model, and the results are below (Model 5)							
Model 5: 1. Dominant arm 2. Modified job duty 3. Age 4. Constant	21.5 (10.6 to 32.3) -22.4 (-37.3 to -7.5) -.96 (-1.52 to -.40) 67.6	.46 -.37 -.42	<.001 .004 .001	.50	.46	-.02	.19
Adding the final non-modifiable factor, a specific cause of the shoulder pain, did not result in any change in the model's R^2 value. Potentially modifiable predictors were assessed next, but neither potential predictor produced a significant change in the R^2 value. Therefore, Model 5 is the most parsimonious model containing only significant predictors of the WORC change score. Below is a full model with all significant univariate predictors from Table 18 included in the Model							
Full Model: Contains all significant univariate predictors from Table 18	na	na		.54	.44	na	na

3.8 FIGURES

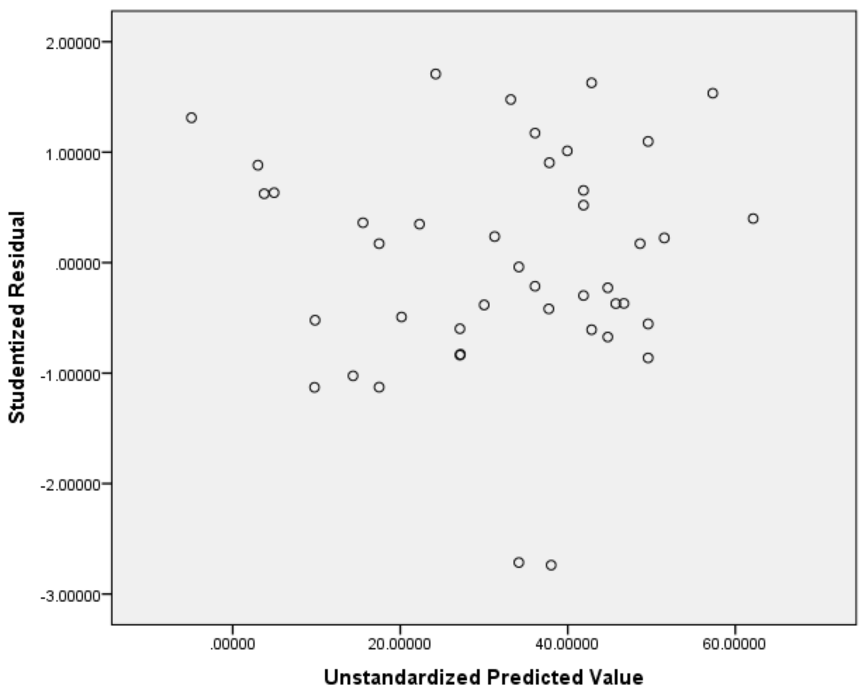


Figure 2 Scatterplot of Studentized Residual against Predicted Values of WORC change score

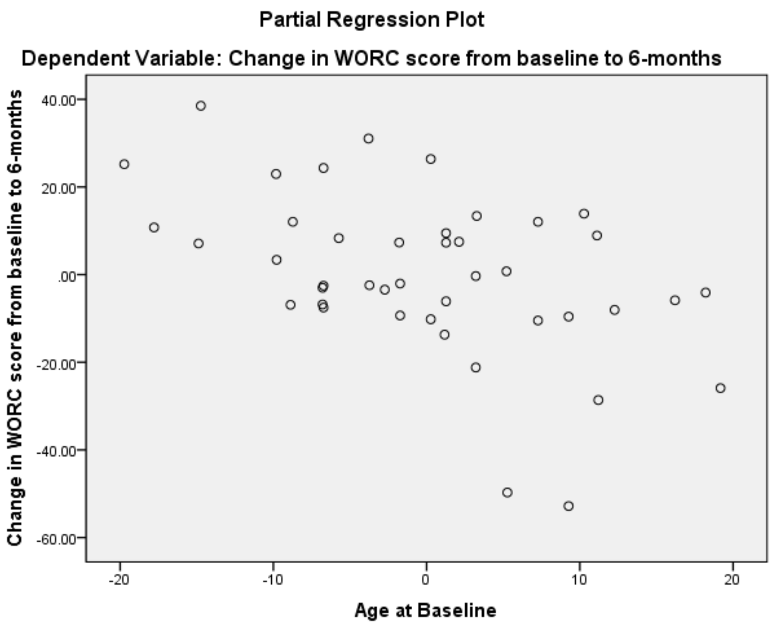


Figure 3 Partial Regression Plot of Age to WORC change score

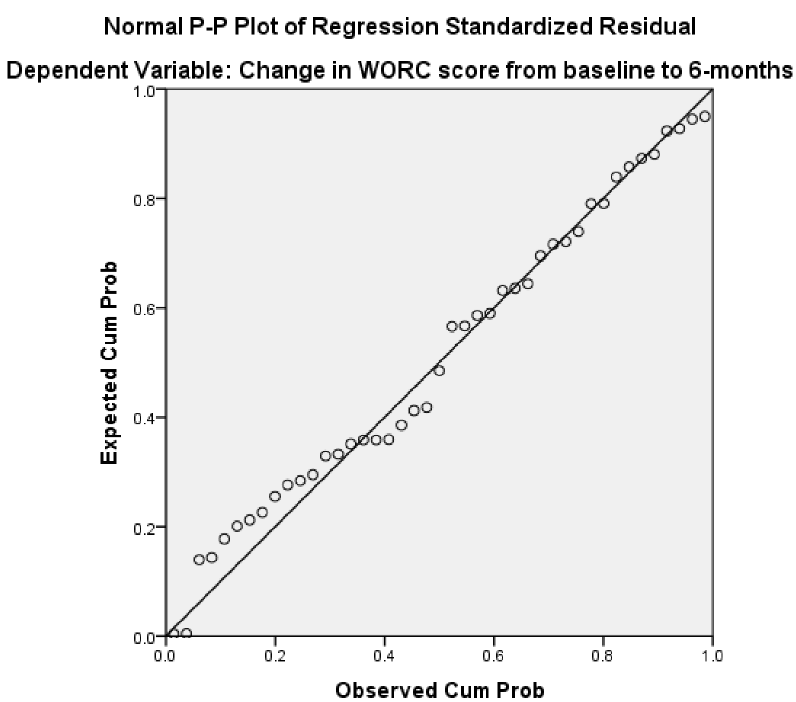


Figure 4 Normal P-P Plot of Regression Standardized Residuals

4.0 RESPONSIVENESS OF THE WESTERN ONTARIO ROTATOR CUFF INDEX AND THE DISABILITIES OF THE ARM, SHOULDER, AND HAND QUESTIONNAIRE IN INDIVIDUALS UNDERGOING ARTHROSCOPIC SURGERY FOR ROTATOR CUFF PATHOLOGY

4.1 INTRODUCTION

With the expected rise in rotator cuff lesions due to aging of the American population, it is important to have clinical outcome measures that capture the patient's perception of their health status following interventions for rotator cuff pathology.[57] Many commonly used instruments for assessing improvement in patients undergoing interventions for rotator cuff pathology lack a combination of proper development, psychometric testing, and/or the ability to capture the multi-dimensional nature of health as defined by the World Health Organization.[20] Two more recently developed patient-reported outcome (PRO) measures to determine change in this population of patients are the Western Ontario Rotator Cuff Index (WORC), a disease-specific PRO measure, and the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH), a region-specific PRO measure.[20, 24] Both of these PROs measure the multi-dimensional nature of health-related quality of life from the patient's perspective and both have been shown to be reliable and valid for use in patients with rotator cuff pathology.[20, 24, 27] To have confidence in the ability of a PRO to detect change from baseline to follow-up, the PRO measure must exhibit both reliability and responsiveness. Reliability is a prerequisite of responsiveness so that there is confidence that an observed change in the PRO measure is due to a true change in the patient's condition, and not simply due to a lack of reliability of the PRO measure.[117] Once determined to be reliable, a PRO measure can then be confidently assessed for its responsiveness, defined as its ability to detect change when a true change has occurred. Although limited, research investigating the responsiveness of the WORC and DASH for surgical treatment of rotator cuff pathology has found both PROs to demonstrate internal responsiveness.[26, 53, 118] Currently, we are unaware of any studies that have reported a MCID for the WORC in individuals undergoing rotator cuff surgery. Gummesson et al[118], in individuals who underwent arthroscopic subacromial decompression surgery, reported a minimally important difference of 10-points on the DASH for individuals reporting they were "somewhat better" and 19-points for those reporting they were "quite a bit better" on a global rating of change. The method used in their study to calculate the minimally important difference was the mean of the change scores for each response level, thus the sensitivity and specificity for the minimum clinically important difference was not determined.

Concurrently investigating the responsiveness of the WORC and the DASH will aid researchers and clinicians in choosing the PRO that most accurately detects changes in the patient's perception of symptoms, activity limitations and participation restrictions in individuals undergoing surgery for rotator cuff pathology.

Measures of responsiveness can be classified as either internal or external measures of responsiveness. According to Husted et al[119], internal responsiveness is the ability of a PRO measure to "detect *change* over a particular specified time frame." Methods for measuring internal responsiveness include effect size, standardized response mean (SRM), standard error of the measure, and the minimal detectable change.[119-121] Measures of internal responsiveness are valuable for researchers to determine required sample sizes for intervention studies in a specific population of patients. However, assessment of internal responsiveness does not allow for determination of a threshold for whether or not individual change has occurred.[120] As such, the clinical utility of internal responsiveness is limited

In contrast to internal responsiveness, external responsiveness can be used to determine when individual change has occurred. According to Husted et al, "External responsiveness reflects the extent to which change in a measure relates to corresponding change in a reference measure of change in clinical or health status."[119] The global rating of change (GRC) is often used as the reference measurement, or anchor, that determines from the patient's perspective, if change in their condition has occurred.[90] By having data that captures change in the WORC or DASH as well as the GRC, it is then possible to calculate a minimally clinically important difference (MCID), which clinicians can use as a threshold to determine when a patient has improved (versus not improved) following treatment (individual change).[90]

The use of a receiver operating characteristic (ROC) curve allows the responsiveness of the WORC and DASH to be expressed and analyzed in terms of sensitivity and specificity of change.[122] Sensitivity of change represents the probability that a cut point in the change score of the WORC or DASH can correctly classify a patient as improved when the external criterion (global rating of change) has indicated an improvement in the condition. Specificity of change is the proportion of individuals who indicated they had not improved on the external criterion who also fell below the cut point in the change score. The area under the ROC curve (AUC) indicates the probability of the WORC or DASH of correctly classifying an individual as improved or not improved for a given cut-point in each instrument's change score.[119] Through ROC curve analysis, it is the MCID that is being determined with the MCID being the smallest change in a PRO measure (i.e. DASH or WORC) that is considered important by the patient. [120]

To date, there have been few studies that have compared the responsiveness of the WORC and DASH simultaneously in patients with rotator cuff pathology who required surgical intervention.[26, 53] In addition, these studies reported only the internal responsiveness of the WORC and DASH and not the external responsiveness (MCID). These studies also varied significantly from the current sample due to the severity of rotator cuff pathology and the interventions used. In the study by MacDermid et al[53], open and mini-open surgical techniques were predominantly used to repair the torn rotator cuff tendon and nearly half of the tears were greater than 3 centimeters. In the study by Lopes et al[26], only a third of their subjects (n=9) underwent a subacromial decompression or rotator cuff repair and the remaining 21 subjects had physical therapy to treat their rotator cuff condition. Therefore, the purpose of this study is to determine the internal and external responsiveness of the WORC and DASH during the first 6 months after surgery in individuals that underwent arthroscopic subacromial decompression with or without arthroscopic repair of small cuff tears.

4.2 METHODS

4.2.1 Subjects

Patients that participated in this study were recruited from the population of patients undergoing arthroscopic subacromial decompression with or without arthroscopic repair of a small rotator cuff tear at the University of Pittsburgh Medical Center (UPMC) Center for Sports Medicine.

The Institutional Review Board at the University of Pittsburgh approved the study, all patients provided informed consent, and the rights of the individuals enrolled as subjects were protected at all times.

Patients scheduled for arthroscopic subacromial decompression with or without a partial or full thickness rotator cuff tear no greater than 2 cm were enrolled in the study. In addition, to participate in the study, individuals had to be 25 years of age or older and have a primary complaint of shoulder pain for at least 3 months, a positive Kennedy-Hawkins or Neer's impingement test and a pain-free contralateral shoulder. The exclusion criteria were: 1) prior glenohumeral dislocation that required relocation by a physician; 2) prior shoulder surgery on the affected side; 3) rheumatoid arthritis; 4) active cervical radiculopathy or 5) expected Grade II or worse labral tear that required repair at the time of surgery.

4.2.2 Procedures

After the patient had met the criteria to participate in the study and informed consent was obtained, a baseline examination was completed and the subjects completed the DASH and WORC questionnaires. Patients then underwent arthroscopic subacromial decompression with repair of the rotator cuff as needed. During the shoulder surgery, if other surgical procedures were needed, such as labral debridement, labral repair, or biceps tenodesis, these procedures were performed and recorded in the operative note. At the 6-month postoperative time point, each subject was mailed the DASH, WORC and GRC, which were returned via mail.

4.2.3 Outcome Measures

The WORC is a disease-specific, PRO that is designed to measure "health-related quality of life" in persons with injuries and conditions of the rotator cuff.[20] The authors, Kirkley et al, intended the measure to represent the impact of rotator cuff disease on health as defined by the World Health Organization – "a state of complete physical, mental and social well-being". The WORC consists of items in 5 domains: 1) pain and physical symptoms; 2) sports and recreation; 3) work; 4) lifestyle and 5) emotions. The questionnaire includes 21 items that are each answered on 100 mm visual analog scale with anchors such as no pain/difficulty and extreme pain/difficulty. For each item, the patient can receive a minimum score of zero (no problem or loss of ability) to 100 (maximum pain/difficulty/disability). The scores for each item are summed and the resulting score is transformed to a scale that ranges from 0 to 100 by subtracting the raw score from 2100, dividing by 2100 and multiplying by 100. A score of 100 indicates no pain/limitation and a score of 0 indicates maximum pain/limitation of the shoulder. The reliability and internal responsiveness of the WORC has been established in earlier studies.[25, 27]

The DASH is a region-specific, 30-item questionnaire that evaluates symptoms and physical function at the level of activity and participation.[24] The DASH is appropriate for measuring symptoms and physical function in patients with any or multiple disorders of the upper limb.[22] Each item of the DASH has five response options used to create a summative score ranging from 0 (no disability or symptoms) to 100 (greatest disability or symptoms). A higher score on the DASH reflects greater disability and worse functional state. A study by Beaton et al[22], in a population of 200 subjects with a variety of upper extremity disorders, found the DASH to have strong construct validity, reliability and responsiveness.

The global rating of change is a 15-point scale that ranges from -7 (a very great deal worse) to +7 (a very great deal better) with 0 in the middle (about the same). Scores of +4 and +5 are indicative of moderate changes in status while +6 and +7 are indicative of a large change in a person's status.[90] The global rating of change has been supported for use as an outcome that measures the patient's perspective as to whether or not they are better.[91] The global rating of change was used to dichotomize patients into those who reported an improvement (successful outcome or responder) and those who did not improve (unsuccessful outcome or nonresponder). To be classified as a responder, the patient had to report a GRC score of +5 or better. This coincides to a self-reported level of improvement from before treatment to 6-months postoperative of "quite a bit better". Any subject reporting that he was at best, "moderately better," (which coincides with a GRC score of +4 or less) was classified as a nonresponder. While this is a stringent criterion for being classified as a responder, since the intervention was surgery, the criterion for success was kept stringent.

4.2.4 Data Analysis

Descriptive statistics were calculated to allow for a comparison between responders and nonresponders. The effect size and SRM were calculated as measures of internal responsiveness.[123-125] ROC curves were constructed to assess the external responsiveness of the WORC and DASH using SPPS version 18 (IBM Corporation Armonk, NY) [119] The criterion for being classified as a responder for the ROC curve analyses was a global rating of change score of +5 or better ("quite a bit better") at follow-up. The ROC curve is created by plotting the sensitivity, or true positive rate on the y-axis and 1-specificity, or the false positive rate on the x-axis for all the WORC and DASH change scores.[122] From the ROC curve analyses, the AUC and the optimal cut-point in the WORC and DASH change scores for dichotomizing subjects into the responder and nonresponder groups was determined. The AUC ranges from 0 to 1.0 and an AUC of 1.0 indicates a perfect ability of the scale to correctly distinguish responders from nonresponders. Selecting the best cut-point to use as the MCID for the WORC and DASH change scores was done using the Youden Index.[94] The Youden Index "reflects the intention of maximizing overall correct classification rates and thus minimizing misclassification rates."[94]

4.3 RESULTS

Sixty-two eligible patients completed baseline testing and 44 patients completed the 6-month follow-up surveys. Comparisons of the 18 subjects who did not complete the 6-month follow-up to those who completed the study are presented in the appendix. In general, those with incomplete follow-up tended to have worse baseline WORC and DASH scores as well as worse depression scores on the Center for Epidemiologic Studies Depression Scale (CES-D). Descriptive statistics for the group that completed the study are presented in Table 20. Based on the GRC, 38 patients (84%) were classified as responders and 6 patients (16%) were classified as nonresponders. Figure 5 presents a scatterplot of the change in WORC to the change in DASH. The correlation between the change scores was 0.66 ($p<.00$). Figure 6 and Table 21 present the change in the WORC and DASH scores by level of global rating of change.

Internal and external responsiveness measures for the whole group and for the responders and nonresponders are presented in Table 22 and Table 23. The effect sizes for the responders on the WORC and DASH were 1.94 and 1.68 respectively, indicating a large effect size for both instruments. Similarly, the SRMs for the responders were 1.67 and 1.58 for the WORC and DASH respectively, which are also considered large.[126]

Figure 7 shows the ROC curves for the WORC and DASH change scores. The AUC for the WORC was 0.84 (95% CI 0.70, 0.98) and the AUC for the DASH was 0.89 (95% CI 0.80.0.99). Both of the AUC values indicate that the WORC and the DASH have excellent capabilities to discriminate between those who perceived themselves to be improved versus those that did not perceive improvement.[127, 128] For the WORC, a MCID of 0 had a sensitivity of change of 0.97and specificity change of 0.50. The DASH had a MCID of -15.0, which had a sensitivity of change of .84 and specificity of change of 1.0. Both of these cut points were chosen to maximize the Youden Index.[94] Using a cut point of 0 for the WORC's MCID resulted in 40 of 44 (91%) subjects being classified correctly as a responder or nonresponder. For the DASH, its MCID resulted in 38 of 44 subjects (86%) being classified correctly.

4.4 DISCUSSION

Large effect sizes and large SRMs were observed for both the WORC and the DASH. For clinicians, the measures of internal responsiveness provide levels of change that should be expected in their patients who are similar to those enrolled in this study. The measures of internal responsiveness also allow for a direct comparison of the WORC to the DASH in determining which instrument more readily captures change when it has occurred, thereby aiding clinicians in choosing the instrument most likely to capture change in their patients. While the WORC had higher values for the internal responsiveness measures, both the WORC and DASH demonstrated large effect sizes and standardized response means indicating both are appropriate for measuring change in this patient population. Therefore, utilizing either the DASH or the WORC would be acceptable and choosing between the two instruments could be based on clinician preference. For researchers, when using a PRO as the primary outcome measure of interest, larger effect sizes and SRMs are beneficial due to the fact that a more responsive outcome measure will require a smaller sample size than would a less responsive outcome measure.[129]

In this study, the WORC demonstrated slightly greater internal responsiveness than the DASH, as measured by the effect size and the SRM (Table 21). MacDermid et al reported slightly larger SRMs at 6-month follow-up for the WORC and DASH in subjects who had rotator cuff repair.[53] Their study reported SRMs of 2.0 and 1.6 for the WORC and DASH, respectively in 86 subjects determined to have a positive response to cuff repair. While they report a slightly larger SRM for both the DASH and WORC, approximately 50% of their subjects had rotator cuff tears greater than 3 cm in size and a majority of the surgeries were open or mini-open procedures. Due to the greater severity of rotator cuff tear in the study by MacDermid et al, their subjects' mean baseline WORC and DASH scores were approximately 30 and 51 respectively for the responder group and improved to 65 on the WORC and 26 on the DASH at the 6-month follow-up. In the current study, the severity of rotator cuff tears was limited to 2 cm by the exclusion criteria and this decreased severity of rotator cuff tears in our study may have resulted in better WORC and DASH scores at baseline (WORC = 44, DASH = 40) and at 6-month follow-up (WORC = 80, DASH = 13) in the group that was considered to be "responders." In addition, all subjects in the current study underwent only arthroscopic procedures. The results reported by MacDermid et al may be more applicable to patients with larger rotator cuff tears who have an open repair while the current study generalizes to patients undergoing arthroscopic subacromial decompression with or without arthroscopic repair of small cuff tears. The contrast of the ES and SRM in our study with those reported by MacDermid et al highlights the dependency of these statistics on the context of the patient population and the type of treatment that was investigated. As such, we believe that the usefulness of these statistics as a measure of responsiveness is somewhat limited.

Another drawback of internal responsiveness is that it has limited clinical utility due to the fact that the calculations are the result of compiling group data without regards to an external criterion for whether improvement in the patient's condition has or has not occurred.[130] Therefore, calculation of the MCID, using each patient's perception of change (GRC), provides clinicians with the magnitude of change of the WORC and DASH which is indicative of self-perceived improvement.[131]

Since there are limited studies that have examined the MCID for the WORC, a main purpose of this study was to collect data that would allow for calculation of the WORC's MCID as well as the MCID for the DASH. However, calculation of a meaningful MCID for the WORC and DASH were not possible because of the large discrepancy between those subjects classified as a responder compared to those classified as nonresponders. While 38 subjects were responders, only 6 were nonresponders. Below, we discuss the small amount of literature that has been published regarding MCIDs for the WORC and DASH in shoulder populations. For the sake of completing the analyses required to calculate the MCIDs, these analyses were completed but the results should be considered invalid. Following the discussion of the current literature on MCIDs for the WORC and DASH, a more detailed explanation of the limitations that resulted from the imbalanced responder/nonresponder rate is provided.

Until now, only 2 studies have calculated the MCID for the WORC,[25, 132] and neither included arthroscopic surgery as the intervention but rather injections for subacromial impingement syndrome. In a study by Kirkley et al,[25] a 17.6 point improvement on the WORC corresponded to "moderate" self-reported improvement of the individual's shoulder condition. Their calculation of the MCID in the subjects who rated their improvement as "moderate" was done using the average change in the WORC score for those individuals and not using an ROC curve analysis. In our study, subjects who reported a GRC score of "somewhat better" also had a mean WORC change score of 17 points. In Ekeberg et al[132], using ROC curve analysis, the MCID for the WORC was 12.8 points at a 6-week follow-up following corticosteroid injection. In that study, a global rating of change scale ranging from -9 (worst change possible) to +9 (best change possible) was used to determine the criterion for improvement. A change of +3 or greater was used as the criterion for improvement in the shoulder condition but there was no verbal anchor to indicate what level of improvement was meant by a +3 score. As with the study by Kirkley et al, the variability in the criterion used for improvement in the shoulder condition and the difference in intervention limits comparisons of those MCIDs to the MCID calculated in this study. It does illustrate the variability in the criterion used to determine the change in a condition and one must be cognizant of this when applying a MCID to a specific patient.

For the DASH, Gummesson et al reported a mean change of either -19 points or -10 points to be the MCID for improvement when the patient reported they were at least "much better" and "somewhat better" respectively following arthroscopic subacromial decompression.[118] In this study, the MCID was calculated as an average of the change score rather than through the use of ROC curve analysis. Schmitt et al[133] also reported a mean change of -21 points and a MCID of -10.2 points on the DASH in individuals treated with physical or occupational therapy. In that study, the criterion for having improved was based on improving at least 1 point on their 7-point global disability scale and the MCID was calculated as the mean change score on the DASH in subjects who indicated at least a 1 point improvement on the global disability scale. The global disability rating used in the study asked the subject to indicate the effect of the injury on their daily function over the past week. The worst score, a 7, indicated maximum disability while the best score, a 1, indicated no disability. The differences in the shoulder pathologies of the study populations and the differences in the criterion used to dichotomize those who have improved from those who have not makes comparisons between the studies difficult.

In the current study, the MCID for the DASH was calculated to be -15.0 points based on ROC curve analysis that utilized the Youden Index to choose the optimal MCID. Again, the MCID for the DASH is suspect given the disproportional balance between responders and nonresponders and the MCID is only given to complete the analyses. If the MCID had been valid, a clinician treating a patient similar to the individuals enrolled in the current study could be confident that a patient with a 15-point improvement on the DASH approximately 6 months after surgery would report at least being “quite a bit better”. The high levels of sensitivity (0.84) and specificity (1.0) for the DASH MCID indicates that classifying a patient based on the DASH’s MCID would result in a correct classification as either a responder or non-responder 86% of the time.

If the MCID for the WORC would have been valid, then a WORC change score of 0 combined with the high sensitivity (0.97) of the cut-point allows the clinician to confidently interpret a patients’ improvement of less than zero points on the WORC as coinciding with a patient’s self-perception of not being improved.

4.5 LIMITATIONS

The major limitation of this study occurred due to the large imbalance of subjects that were classified by the GRC as responders (38 of 44 subjects) as compared to nonresponders (6 of 44 subjects). Due to this imbalance, maximizing the sensitivity, rather than the specificity, of the MCID cut point was necessary to maximize the correct classification rate which is the goal of the Youden index. Specificity, in the end, carried much less influence on the classification accuracy. Thus, a specificity of 0.50 for the WORC MCID cut point was allowable without significantly impacting the overall classification accuracy. The lack of impact of specificity on the MCID cut-point combined with only a handful of subjects that did not have a large improvement on both the WORC and DASH resulted in a nonsensical WORC MCID of zero.

An appropriate strategy to increase the percentage of nonresponders would be to recruit non-operative shoulder pain patients who are less likely to respond to conservative treatment. By including more subjects who are likely to be classified as nonresponders, the specificity of the MCID cut point will have a greater impact on overall classification accuracy and produce a MCID for the WORC and DASH that is calculated from a more balance ratio of responders to nonresponders.

While the GRC was used as the criterion on which the MCID was calculated, the GRC does lack reliability testing and additional techniques to improve recall of the patient's condition prior to treatment were not used in this study.[134]

Finally, it has been shown that varying methods for determining when change has occurred can result in very different proportions of the same patient population being classified as "responder" versus "nonresponder" For instance, Beaton et al utilized 13 various, but published, approaches to determining the MCID and found that the proportion of the subjects classified as a responder varied from 39 to 89 percent depending on the approach.[135] More recently, the use of the patient acceptable symptom state (PASS) as a criterion for determining successful interventions has gained popularity.[136-138] The PASS divides patients into groups that are either satisfied with their state of health following an intervention or not satisfied. To determine if they were satisfied or not, patients were asked, "taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?''[136] The PASS is seen as a complimentary measure to the MCID in that the MCID focus on change over treatment while the PASS focuses on only the final state. Determining both the MCID and PASS would allow results of studies to be presented as proportions of patients that improved as well as the proportion that report being in a satisfied state following treatment.

4.6 CONCLUSION

In light of the limitations, this study provides evidence that both the WORC and the DASH exhibit high levels of internal responsiveness. However, give the limitations described above, it does not provide sufficient evidence for the MCID of the WORC or DASH.

4.7 TABLES

Table 20 Baseline characteristics (n=44)

Variable	Mean (SD) or Frequency (%)
Age, years	46.1 (10.3)
Female, n (%)	28 (64%)
Duration of Pain, months (%)	
3-6	13 (31%)
7-12	9 (21%)
13-24	13 (31%)
≥ 25	7 (17%)
Surgical Procedure (structures repaired)ᵻ	
ASD only	9 (20%)
Supraspinatus	7 (16%)
Supraspinatus, labrum	1 (2%)
Supraspinatus, biceps	2 (5%)
Supraspinatus, labrum, biceps	1 (2%)
Infraspinatus	2 (5%)
Infraspinatus, labrum	1 (2%)
Subscapularis	10 (23%)
Subscapularis, labrum	2 (5%)
Subscapularis, biceps	1 (2%)
Labrum	3 (7%)
Labrum, biceps	2 (5%)
Subscapularis, supraspinatus	3 (7%)
Baseline WORC score	43.8 (19.5)
Baseline DASH score	40.0 (17.3)

ᵻAll procedures were in addition to arthroscopic subacromial decompression (ASD)

Table 21 Change in WORC and DASH scores by Global rating of change

Global rating of change at 6-months postop	Change in WORC Mean ± SD (range)	Change in DASH Mean ± SD (range)
+7 a very great deal better (n=13)	37 ± 22(0 to 69)	-26 ± 11 (-42 to -6)
+6 a great deal better (n=14)	40 ± 21 (-10 to 82)	-25 ± 15 (-51 to 2)
+5 quite a bit better (n=10)	30 ± 23 (2 to 69)	-29 ± 26 (-61 to 20)
+3 somewhat better (n=2)	17 ± 4 (15 to 20)	-8 ± 7 (-13 to -3)
+2 a little bit better (n=2)	12 ± 27 (-7 to 31)	-3 ± 16 (-14 to 8)
-5 quite a bit worse (n=1)	-8	-3
-6 a great deal worse (n=1)	-1	-1

Table 22 Mean changes by group and internal responsiveness of the WORC and DASH

	WORC					DASH				
	baseline	6m	Δ	ES	SRM	baseline	6m	Δ	ES	SRM
Group means ±SD										
All subjects (n=44)	44±20	77±24	32±23	1.63	1.41	40±17	16±19	-24±18	1.36	1.33
GRC ≥ +5 group* (n=38)	44±18	80±22	36±21	1.94	1.67	40±16	13±15	-27±17	1.68	1.58
GRC ≤ +4 group (n=6)	43±28	51±28	8±16	0.30	0.52	39±27	35±30	-4±8	0.16	0.51

*A global rating of change score of +5 or greater indicates the patient reported they were at the least "quite a bit better" at the time the survey was administered compared to their condition prior to shoulder surgery. A global rating of change score of +4 or less indicates the patient reported they were at best "moderately better" Abbreviations: ES, effect size; SRM; standardized response mean; WORC, Western Ontario Rotator Cuff Index; DASH, Disabilities of the Arm, Shoulder and Hand; GRC, global rating of change; MCID, minimal clinically important difference; Δ, change

Table 23 Minimal Clinically Important Difference of the WORC and DASH

	WORC	DASH
MCID	0	-15
Sn/Sp of MCID cut point	.97/.50	.84/1.0
Area under the ROC curve (95%CI)	0.84 (0.70, 0.98)	0.89 (0.80, 0.99)

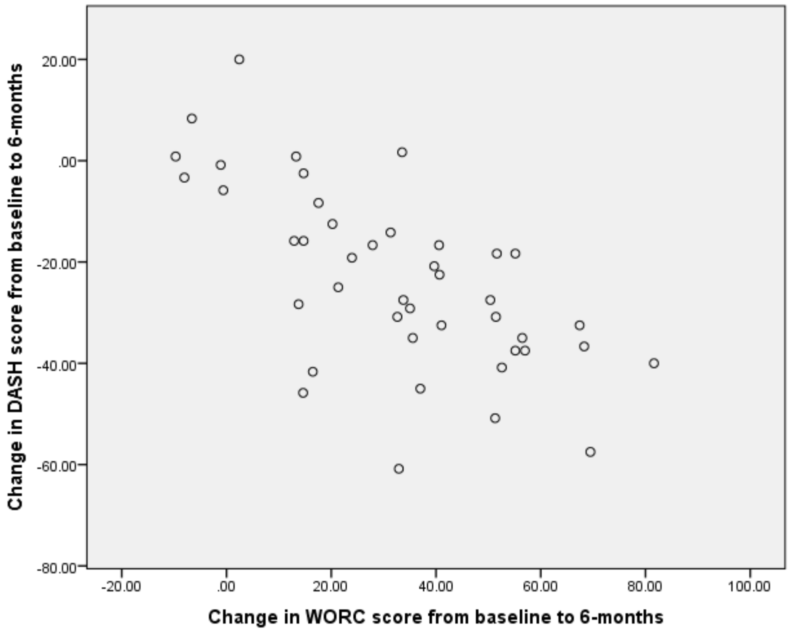
Abbreviations: Sn, sensitivity; Sp, specificity

Table 24 Comparison of subjects with complete and incomplete follow-up at 6-months*

	Complete data (n=44)	Incomplete data (n=18)	p-value (Mann Whitney U)
Age	46±10	46±13	.91
WORC baseline	44±19	33±19	.04[ᵻ]
DASH baseline	40±17	47±15	.09
FABQ_work subscale	16±15	23±13	.10
CES_D	6±7	12±11	.03[ᵻ]

*Values are mean ± SD, [ᵻ] p-value < 0.05

4.8 FIGURES



r = 0.66, p<.00

Figure 5 Scatterplot of change in DASH by change in WORC

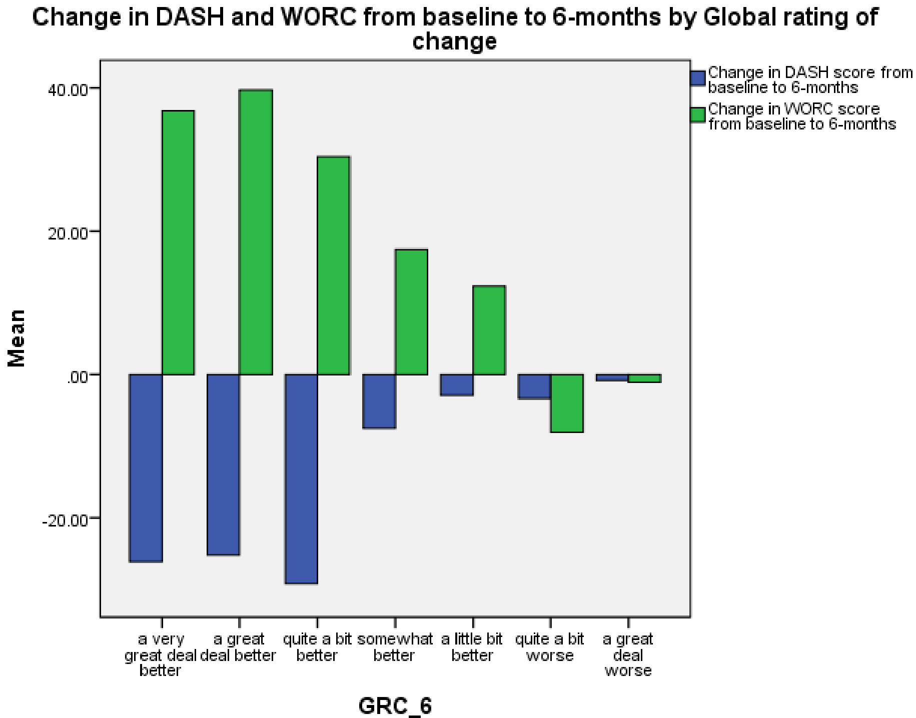


Figure 6 Mean change in DASH and WORC by global rating of change score at 6-months

Abbreviations: GRC_6, global rating of change score at 6-month follow-up

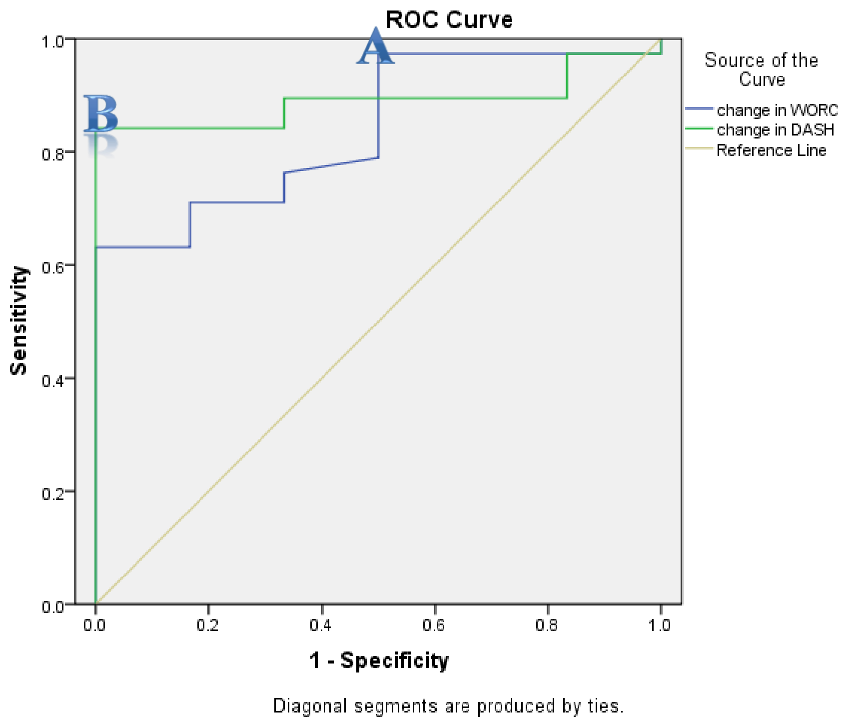


Figure 7 ROC curves for WORC and DASH change scores

Legend: A, indicates on the ROC curve the combination of sensitivity and specificity for the cut-point in the WORC change score that maximized correct classification of responders to nonresponders; B, indicates on the ROC curve the combination of sensitivity and specificity for the cut-point in the DASH change score that maximized correct classification of responders to nonresponders

5.0 SIGNIFICANCE AND DIRECTION OF FUTURE RESEARCH

As part of evidence-based medicine, care providers should utilize prognostic evidence to help patients make treatment decisions by discussing with the patient their probability of successful outcome following an intervention. Research studies designed to evaluate predictors of rotator cuff surgical outcome, and therefore to provide prognostic evidence, should attempt to provide high levels of evidence by utilizing a strong study design, logical candidate predictors, appropriate statistical analyses, and valid, responsive patient-reported outcomes.[56] Once prognostic studies identify factors that impact outcome, these results can be used to not only inform clinicians and their patients regarding prognosis, but also to develop targeted interventions for the potentially modifiable factors that impact outcome.

In individuals that have rotator cuff repair, our systematic review of this literature revealed equivocal results for the prognostic ability of many of the examined predictors. The prognostic factor with the least equivocal evidence for impacting clinical outcome was a worker's compensation claim that pertained to the rotator cuff tear. Also based on our systematic review, increasing age appears to predict worse outcomes when the DASH is used as the dependent outcome. Based on the results of our systematic review, other potential prognostic factors did not demonstrate an ability to consistently predict outcome in this population. Sex, fatty infiltration of the cuff musculature, smoking status, comorbidities, involvement of the dominant arm, and duration of symptoms, were potential predictors that were examined, and all failed to show any consistent relationship with clinical outcome.

In addition to demonstrating a lack of strong and consistent predictors of clinical outcome, our systematic review also sheds light on the myriad of outcome measures used in this body of research. The large variety of outcome measures used results in difficulty comparing the findings of these prognostic studies. Instruments such as the Constant score and the UCLA Shoulder score frequently appear as the outcome measures in the studies included in our systematic review. Both of these instrument lack modern development, psychometric testing, and rely largely on clinician ratings of impairments such as shoulder range of motion and strength to contribute to the final score. The inclusion of these clinician measured items eliminates both the Constant score and the UCLA Shoulder score from being purely patient-reported outcome instruments.

Due to the weaknesses found in many of the commonly used outcome measures, our study, which examined the predictors of patient-reported outcome following surgery for rotator cuff pathology, chose to use the WORC and the global rating of change as the outcome measures that would determine responders from nonresponders. To provide a comparison to the logistic model, a linear regression model was also built using the WORC change score as the dependent variable. The WORC is an outcome measure that was developed using modern methods, incorporates the multi-dimensional nature of health, is purely patient-reported, and has acceptable psychometric properties.[20, 26, 27, 139] The global rating of change provided another patient-reported outcome measure that indicates the patient's perception of their improvement.[90, 91] The stringent criteria used in our study to determine a responder (17 point improvement on the WORC and a +5 ("quite a bit better") or better on the global rating of change undoubtedly lowered our responder rate. However, due to the stringent criteria for achieving responder status, an individual with a high predicted probability of success, based on the results of our logistic regression analyses, can be confident that they will have a successful surgical outcome.

We also chose to include as candidate predictors commonly used tests and measures that are readily collected during a routine clinical examination. For example, for measures of body structure and function, we included candidate predictors that quantified scapular dyskinesis, shoulder strength using handheld dynamometry, and internal rotation range of motion of the glenohumeral joint. Measures such as these are commonly performed during an examination of these patients and including them in our study allowed us to examine their prognostic ability. Beyond their prognostic ability, impairments of body structure and function are modifiable factors that can be treated in physical therapy before and/or after surgery. While internal rotation strength as a percentage of the uninvolved side was a univariate predictor of response to surgery, it did not remain an independent predictor of outcome once other predictors that demonstrated a significant univariate relationship with response were also included in the regression model. Due to the smaller sample size of our study, the inclusion of internal rotation strength deficits in future prognostic studies should be considered and explored further for its relationship to successful surgery in this patient population. If shown to be a predictor, then further studies would be warranted to see if it is possible to improve this deficit and if doing so improves patient-reported outcomes 6 months after surgery.

In concordance with the results of our systematic review, we also found worker's compensation status to be a factor that had a significant univariate relationship for decreasing the probability of being a responder to surgery. However, it did not remain an independent predictor once the FABQ_work subscale was included in the multivariable logistic regression analysis.

The FABQ_work scale, on the other hand, did remain a significant predictor when worker's compensation status was also included as a predictor in the multivariable logistic regression analysis. In this study, the FABQ_work subscale and worker's compensation status were moderately to strongly correlated. This resulted in the variables competing for entry into the regression equations which was not the case in previous studies that did not include measures of fear-avoidance as potential predictors of outcome in this population of patients.

Of the two predictors that remained in the final logistic regression model, the FABQ_work subscale and having surgery on the dominant arm, the presence of high fear-avoidance beliefs provides a potentially modifiable factor that might be managed through the use of graded exercise activities and patient education.[105-107] In patients with low back pain, a fear-avoidance based treatment protocol has been shown to improve outcomes in those with high fear-avoidance levels.[105] If FABQ scores are consistently shown to be a predictor of outcome in this population, then efforts should be made to determine if addressing high levels of fear-avoidance with specific fear-avoidance treatment protocols can result in improved outcomes in patients undergoing surgery for rotator cuff pathology.

The linear regression models also found that surgery on the dominant arm was a predictor of the WORC change score. Therefore, both models had surgery on the dominant arm as an independent predictor of outcome. While the FABQ_work score was not significant in the linear model it was a univariate predictor of the WORC change score. It may have been excluded from the final model due to the fact that the variable modified job duty was moderately correlated to the FABQ scores and the subjects who were on modified job duty had high FABQ_work subscale scores while those subjects not on modified job duty tended to have much lower FABQ_work subscale scores (less fear-avoidance beliefs). Therefore, once in the model, modified job duty explained much of the variance that would also be explained by the fear-avoidance score. Age, was a predictor of the WORC change score, with increasing age resulting in less improvement on the WORC. Age was not a predictor in the logistic model. These findings are reflective of what was discovered in our systematic review. Age may be a predictor of somewhat worse outcomes following rotator cuff surgery, but its ability to be a predictor is not consistent and may potentially be outcome and sample specific.

Finally, our study aimed to determine the MCID for both the WORC and the DASH in this patient population and to provide a comparison of these PRO measures in terms of their internal and external responsiveness. For internal responsiveness, the WORC and the DASH demonstrated large effect sizes and large standardized response means. This indicates that clinicians can be confident that the WORC and the DASH are suitable for capturing change in this population. For external responsiveness, the lack of nonresponders did not allow for a calculation of a sensible MCID value. Therefore, the MCID calculations performed in this dissertation were to complete the analyses but the results should not be considered valid. The results of this dissertation indicate that continued research into the prognostic factors that impact outcome of surgery for the rotator cuff are still needed. The strength of a psychosocial predictor, fear-avoidance, is not unexpected given its prognostic ability in other musculoskeletal disorders. Still, this study brings to light the need for additional investigation of this factor for both its prognostic ability in surgery for rotator cuff pathology and as a modifiable factor that can be treated by specific physical therapy interventions. Finally, to make the prognostic studies comparable, there needs to be adoption of a standard outcome measure that is valid and responsive for assessing change in patient-reported outcome in this population. While multiple outcomes could be used in a study, adoption of 1 standard outcome measure would make the body of prognostic evidence more cohesive and useful. This study provided evidence that both the WORC and the DASH have the potential to serve as that standard PRO measure.

APPENDIX A

TESTS OF SCAPULAR DYSKINESIS

Scapular Assist Test

The modified SAT will be performed as described by Rabin et al.[17] The tester will stand behind the subject and place one hand on the superior aspect of the involved scapula with the finger over the clavicle. The other hand will be placed over the inferior angle of the scapula with the palm of the hand over the inferior angle and the fingers wrapping around the lateral thorax. The subject is asked to elevate his arm in the sagittal plane and the tester provides assistance with upward rotation of the scapula by pushing both upwards and laterally on the inferior angle. At the same time, the tester uses the top hand to tilt the scapula posteriorly by pulling backwards on the superior aspect of the scapula. Pain measures on an 11 point scale will be taken at rest, after the subject actively elevates the arm as high as possible in the sagittal plane without scapular assistance, and after elevating the arm as high as possible in the sagittal plane while the tester provides scapular assistance. A test will be considered positive when the pain rating from elevation of the arm without assistance is 2 points or more greater than the pain rating given by the subject when the arm is elevated with scapular assistance from the tester.

Scapular Dyskinesis System

The SDS will be performed with the subject in standing and the tester behind the subject. The subject will be instructed on elevate the arm in both the frontal plane and in the scapular plane at a rate of 45º per second. Each subject will elevate the arm 3 times in each plane in a counterbalanced order. Abnormal scapular motion/positioning will be categorized into 1 of the 3 patterns described by Kibler

The Scapular Index

To calculate the Scapula Index, the examiner measures from the sternal notch (SN) to the medial aspect of the coracoid process (CP). The examiner then measures the horizontal distance from the posterolateral corner of the acromion (PLA) to the thoracic spine (TS). The SN to CP distance is then divided by the PLA to TS distance and the resulting value is multiplied by 100.

Scapula Index = (SN to CP) / (PLA to TS) x 10

APPENDIX B

THE FEAR AVOIDANCE BELIEFS QUESTIONNAIRE

Here are some of the things other patients have told us about their pain. For each statement please check **one** box to the right of each statement to indicate how much physical activities such as **lifting a heavy object, carrying groceries, reaching overhead, driving a car, or putting on clothes would affect your shoulder pain**

Table 25 Fear Avoidance Beliefs Questionnaire: Shoulder

	Completely Disagree **0**	**1**	**2**	**Unsure** **3**	**4**	**5**	**Completely Agree** **6**
1. My pain was caused by physical activity.	☐	☐	☐	☐	☐	☐	☐
2. Physical activity makes my pain worse.	☐	☐	☐	☐	☐	☐	☐
3. Physical activity might harm my shoulder.	☐	☐	☐	☐	☐	☐	☐
4. I should not do physical activities which (might) make my pain worse.	☐	☐	☐	☐	☐	☐	☐
5. I cannot do physical activities which (might) make my pain worse.	☐	☐	☐	☐	☐	☐	☐

The following statements are about how your normal work affects or would affect your shoulder pain:

Table 25 (continued)

	Completely Disagree 0	1	2	Unsure 3	4	5	Completely Agree 6
6. My pain was caused by my work or by an accident at work.	☐	☐	☐	☐	☐	☐	☐
7. My work aggravated my pain.	☐	☐	☐	☐	☐	☐	☐
8. I have a claim for compensation for my pain.	☐	☐	☐	☐	☐	☐	☐
9. My work is too heavy for me.	☐	☐	☐	☐	☐	☐	☐
10. My work makes or would make my pain worse.	☐	☐	☐	☐	☐	☐	☐
11. My work might harm my shoulder.	☐	☐	☐	☐	☐	☐	☐
12. I should not do my regular work with my present pain	☐	☐	☐	☐	☐	☐	☐
13. I cannot do my normal work with my present pain.	☐	☐	☐	☐	☐	☐	☐
14. I cannot do my normal work until my pain is treated.	☐	☐	☐	☐	☐	☐	☐
15. I do not think that I will be back to my normal work within 3 months.	☐	☐	☐	☐	☐	☐	☐
16. I do not think that I will ever be able to go back to that work.	☐	☐	☐	☐	☐	☐	☐

BIBLIOGRAPHY

1. Balyk R, Luciak-Corea C, Otto D, Baysal D, Beaupre L. Do outcomes differ after rotator cuff repair for patients receiving workers' compensation? *Clin Orthop Relat Res.* Dec 2008;466(12):3025-3033.
2. Baysal D, Balyk R, Otto D, Luciak-Corea C, Beaupre L. Functional outcome and health-related quality of life after surgical repair of full-thickness rotator cuff tear using a mini-open technique. *Am J Sports Med.* Sep 2005;33(9):1346-1355.
3. Boissonnault WG, Badke MB, Wooden MJ, Ekedahl S, Fly K. Patient outcome following rehabilitation for rotator cuff repair surgery: the impact of selected medical comorbidities. *J Orthop.Sports Phys Ther.* 2007;37(6):312-319.
4. Charousset C, Grimberg J, Duranthon LD, Bellaiche L, Petrover D, Kalra K. The time for functional recovery after arthroscopic rotator cuff repair: correlation with tendon healing controlled by computed tomography arthrography. *Arthroscopy.* Jan 2008;24(1):25-33.
5. Davis AM, Beaton DE, Hudak P, et al. Measuring disability of the upper extremity: a rationale supporting the use of a regional outcome measure. *J Hand Ther.* 1999;12(4):269-274.
6. Gartsman GM, Brinker MR, Khan M. Early effectiveness of arthroscopic repair for full-thickness tears of the rotator cuff: an outcome analysis. *J Bone Joint Surg Am.* Jan 1998;80(1):33-40.
7. Goutallier D, Postel JM, Bernageau J, Lavau L, Voisin MC. Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. *Clin Orthop Relat Res.* Jul 1994(304):78-83.
8. Gulotta LV, Nho SJ, Dodson CC, Adler RS, Altchek DW, MacGillivray JD. Prospective evaluation of arthroscopic rotator cuff repairs at 5 years: part I--functional outcomes and radiographic healing rates. *J Shoulder Elbow Surg.* Sep;20(6):934-940.
9. Henn RF, III, Kang L, Tashjian RZ, Green A. Patients' preoperative expectations predict the outcome of rotator cuff repair. *J Bone Joint Surg Am.* 2007;89(9):1913-1919.
10. Hersch JC, Sgaglione NA. Arthroscopically assisted mini-open rotator cuff repairs. Functional outcome at 2- to 7-year follow-up. *Am J Sports Med.* May-Jun 2000;28(3):301-311.
11. Milano G, Grasso A, Salvatore M, Zarelli D, Deriu L, Fabbriciani C. Arthroscopic rotator cuff repair with and without subacromial decompression: a prospective randomized study. *Arthroscopy.* Jan 2007;23(1):81-88.

12. Milano G, Grasso A, Zarelli D, Deriu L, Cillo M, Fabbriciani C. Comparison between single-row and double-row rotator cuff repair: a biomechanical study. *Knee Surg Sports Traumatol Arthrosc.* Jan 2008;16(1):75-80.
13. Lim JT, Acornley A, Dodenhoff RM. Recovery after arthroscopic subacromial decompression: prognostic value of the subacromial injection test. *Arthroscopy.* 2005;21(6):680-683.
14. Oh JH, Kim SH, Ji HM, Jo KH, Bin SW, Gong HS. Prognostic factors affecting anatomic outcome of rotator cuff repair and correlation with functional outcome. *Arthroscopy.* Jan 2009;25(1):30-39.
15. Oh JH, Kim SH, Kang JY, Oh CH, Gong HS. Effect of age on functional and structural outcome after rotator cuff repair. *Am J Sports Med.* Apr 2010;38(4):672-678.
16. Pai VS, Lawson DA. Rotator cuff repair in a district hospital setting: outcomes and analysis of prognostic factors. *J Shoulder Elbow Surg.* May-Jun 2001;10(3):236-241.
17. Patel VR, Singh D, Calvert PT, Bayley JI. Arthroscopic subacromial decompression: results and factors affecting outcome. *J Shoulder.Elbow.Surg.* 1999;8(3):231-237.
18. Watson AW, Mac DC. A reliable technique for the assessment of posture: assessment criteria for aspects of posture. *J Sports Med Phys Fitness.* 2000;40(3):260-270.
19. Preamble to the Constitution of the World Health Organization. *International Health Conference*. New York; 1948.
20. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med.* 2003;13(2):84-92.
21. Atroshi I, Gummesson C, Andersson B, Dahlgren E, Johansson A. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: reliability and validity of the Swedish version evaluated in 176 patients. *Acta Orthop.Scand.* 2000;71(6):613-618.
22. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther.* 2001;14(2):128-146.
23. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum.Dis.* 2004;63(4):335-341.
24. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind.Med.* 1996;29(6):602-608.
25. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy.* 2003;19(10):1109-1120.
26. Lopes AD, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, Dos Reis FB. Validity and reliability of the Western Ontario Rotator Cuff Index (WORC) for use in Brazil. *Clin J Sport Med.* May 2008;18(3):266-272.
27. Holtby R, Razmjou H. Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. *J Shoulder.Elbow.Surg.* 2005;14(5):506-510.
28. Surgeons AAoO. *The Burden of musculoskeletal Diseases*. 1st ed: American Academy of Orthopaedic Surgeons; 2008.

29. Garrett WE, Jr., Swiontkowski MF, Weinstein JN, et al. American Board of Orthopaedic Surgery Practice of the Orthopaedic Surgeon: Part-II, certification examination case mix. *J Bone Joint Surg Am.* 2006;88(3):660-667.

30. Mauro C, Jordan S, Irrgang J, Harner C. Practice Patterns for Subacromial Decompression and Rotator Cuff Repair. *Journal of Bone and Joint Surgery.* 2012.

31. Bjornsson HC, Norlin R, Johansson K, Adolfsson LE. The influence of age, delay of repair, and tendon involvement in acute rotator cuff tears: structural and clinical outcomes after repair of 42 shoulders. *Acta Orthop.* Apr;82(2):187-192.

32. Bohannon RW, Corrigan D. A broad range of forces is encompassed by the maximum manual muscle test grade of five. *Percept.Mot.Skills.* 2000;90(3 Pt 1):747-750.

33. Prasad N, Odumala A, Elias F, Jenkins T. Outcome of open rotator cuff repair. An analysis of risk factors. *Acta Orthop Belg.* Dec 2005;71(6):662-666.

34. Mallon WJ, Misamore G, Snead DS, Denton P. The impact of preoperative smoking habits on the results of rotator cuff repair. *J Shoulder.Elbow.Surg.* 2004;13(2):129-132.

35. Henn RF, 3rd, Kang L, Tashjian RZ, Green A. Patients with workers' compensation claims have worse outcomes after rotator cuff repair. *J Bone Joint Surg Am.* Oct 2008;90(10):2105-2113.

36. McKee MD, Yoo DJ. The effect of surgery for rotator cuff disease on general health status. Results of a prospective trial. *J Bone Joint Surg Am.* Jul 2000;82-A(7):970-979.

37. Namdari S, Baldwin K, Glaser D, Green A. Does obesity affect early outcome of rotator cuff repair? *J Shoulder Elbow Surg.* Dec 2010;19(8):1250-1255.

38. Hayes KW, Petersen CM. Reliability of classifications derived from Cyriax's resisted testing in subjects with painful shoulders and knees. *J Orthop.Sports Phys Ther.* 2003;33(5):235-246.

39. Tashjian RZ, Henn RF, Kang L, Green A. Effect of medical comorbidity on self-assessed pain, function, and general health status after rotator cuff repair. *J Bone Joint Surg Am.* 2006;88(3):536-540.

40. Garrard J. *Health Sciences Literature Review Made Easy: The Matrix Method*. 3rd ed: Jones&Bartlett Learning; 2011.

41. Beattie PF, Nelson RM. Evaluating research studies that address prognosis for patients receiving physical therapy care: a clinical update. *Phys Ther.* Nov 2007;87(11):1527-1535.

42. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med.* Mar 21 2006;144(6):427-437.

43. Hess DR. Retrospective studies and chart reviews. *Respir Care.* Oct 2004;49(10):1171-1174.

44. Portney L, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Stamford: Appleton & Lange; 1993.

45. De Bie R. Critical appraisal of prognostic studies: an introduction. *Physiotherapy Theory and Practice.* 2001;17:161-172.

46. Sackett DL, Straus S, Richardson W. *Evidence-Based Medicine: How to Practice and Teach EBM.* New York: Churchill Livingstone, Inc; 2000.

47. Guyatt G, Rennie D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 2nd ed. Chicago: AMA Press; 2002.

48. AAOS. Optimizing the Management of Rotator Cuff Problems. *American Academy of Orthopaedic Surgeons.* 2010;1st Edition.

49. Jorgensen LN, Kallehave F, Christensen E, Siana JE, Gottrup F. Less collagen production in smokers. *Surgery.* Apr 1998;123(4):450-455.

50. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum.* May 15 2009;61(5):623-632.

51. Amstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. *Clin Orthop Relat Res.* Mar-Apr 1981(155):7-20.

52. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy.* Dec 2003;19(10):1109-1120.

53. MacDermid JC, Drosdowech D, Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. *J Shoulder.Elbow.Surg.* 2006;15(4):407-414.

54. Godfrey J, Hamman R, Lowenstein S, Briggs K, Kocher M. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *J Shoulder Elbow Surg.* May-Jun 2007;16(3):260-267.

55. Organization WH. The Constitution of the World Health Organization. Geneva, Switzerland; 1984.

56. Howick J, Chalmers I, Glasziou P, et al. The Oxford 2011 Levels of Evidence. http://www.cebm.net/index.aspx?o=5653.

57. AAOS. *The Burden of musculoskeletal Diseases*. 1st ed: American Academy of Orthopaedic Surgeons; 2008.

58. Goldman L, Cook EF, Johnson PA, Brand DA, Rouan GW, Lee TH. Prediction of the need for intensive care in patients who come to the emergency departments with acute chest pain. *N Engl J Med.* Jun 6 1996;334(23):1498-1504.

59. Fischer JE, Steiner F, Zucol F, et al. Use of simple heuristics to target macrolide prescription in children with community-acquired pneumonia. *Arch Pediatr Adolesc Med.* Oct 2002;156(10):1005-1008.

60. Liem D, Lichtenberg S, Magosch P, Habermeyer P. Magnetic resonance imaging of arthroscopic supraspinatus tendon repair. *J Bone Joint Surg Am.* Aug 2007;89(8):1770-1776.

61. Harryman DT, Sidles JA, Clark JM, McQuade KJ, Gibb TD, Matsen FA, III. Translation of the humeral head on the glenoid with passive glenohumeral motion. *J Bone Joint Surg Am.* 1990;72(9):1334-1343.

62. Tyler TF, Nicholas SJ, Roy T, Gleim GW. Quantification of posterior capsule tightness and motion loss in patients with shoulder impingement. *Am J Sports Med.* 2000;28(5):668-673.

63. McClure PW, Michener LA, Karduna AR. Shoulder function and 3-dimensional scapular kinematics in people with and without shoulder impingement syndrome. *Phys Ther.* 2006;86(8):1075-1090.

64. Tyler TF, Nahow RC, Nicholas SJ, McHugh MP. Quantifying shoulder rotation weakness in patients with shoulder impingement. *J Shoulder.Elbow.Surg.* 2005;14(6):570-574.

65. Leroux JL, Codine P, Thomas E, Pocholle M, Mailhe D, Blotman F. Isokinetic evaluation of rotational strength in normal shoulders and shoulders with impingement syndrome. *Clin Orthop.Relat Res.* 1994(304):108-115.

66. Graichen H, Stammberger T, Bonel H, et al. Three-dimensional analysis of shoulder girdle and supraspinatus motion patterns in patients with impingement syndrome. *J Orthop.Res.* 2001;19(6):1192-1198.

67. Ludewig PM, Cook TM. Alterations in shoulder kinematics and associated muscle activity in people with symptoms of shoulder impingement. *Phys Ther.* 2000;80(3):276-291.

68. Crombez G, Vlaeyen JW, Heuts PH, Lysens R. Pain-related fear is more disabling than pain itself: evidence on the role of pain-related fear in chronic back pain disability. *Pain.* 1999;80(1-2):329-339.

69. Turk DC, Okifuji A. Psychological factors in chronic pain: evolution and revolution. *J Consult Clin Psychol.* 2002;70(3):678-690.

70. Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine.* 2002;27(5):E109-E120.

71. Ozcetin A, Ataoglu S, Kocer E, et al. Effects of depression and anxiety on quality of life of patients with rheumatoid arthritis, knee osteoarthritis and fibromyalgia syndrome. *West Indian Med J.* 2007;56(2):122-129.

72. George SZ, Dover GC, Fillingim RB. Fear of pain influences outcomes after exercise-induced delayed onset muscle soreness at the shoulder. *Clin J Pain.* 2007;23(1):76-84.

73. Ring D, Kadzielski J, Fabian L, Zurakowski D, Malhotra LR, Jupiter JB. Self-reported upper extremity health status correlates with depression. *J Bone Joint Surg Am.* 2006;88(9):1983-1988.

74. Brophy RH, Beauvais RL, Jones EC, Cordasco FA, Marx RG. Measurement of shoulder activity level. *Clin Orthop.Relat Res.* 2005;439:101-108.

75. Jacob T, Baras M, Zeev A, Epstein L. Low back pain: reliability of a set of pain measurement tools. *Arch Phys Med Rehabil.* 2001;82(6):735-742.

76. Klenerman L, Slade PD, Stanley IM, et al. The prediction of chronicity in patients with an acute attack of low back pain in a general practice setting. *Spine.* 1995;20(4):478-484.

77. Mintken PE, Cleland JA, Whitman JM, George SZ. Psychometric properties of the Fear-Avoidance Beliefs Questionnaire and Tampa Scale of Kinesiophobia in patients with shoulder pain. *Arch Phys Med Rehabil.* Jul;91(7):1128-1136.

78. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement.* 1977;1(3):385-401.

79. Hann D, Winter K, Jacobsen P. Measurement of depressive symptoms in cancer patients: Evaluation of the center for epidemiological studies depression scale (Ces-d). *Journal of Psychosomatic Research.* 1999;46(5):437-443.

80. Kabacoff RI, Segal DL, Hersen M, Van Hasselt VB. Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *J Anxiety.Disord.* 1997;11(1):33-47.

81. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol.* 1988;56(6):893-897.

82. Osman A, Hoffman J, Barrios FX, Kopper BA, Breitenstein JL, Hahn SK. Factor structure, reliability, and validity of the Beck Anxiety Inventory in adolescent psychiatric inpatients. *J Clin Psychol.* 2002;58(4):443-456.

83. Awan R, Smith J, Boon AJ. Measuring shoulder internal rotation range of motion: a comparison of 3 techniques. *Arch Phys Med Rehabil.* 2002;83(9):1229-1234.

84. Kelly BT, Kadrmas WR, Speer KP. The manual muscle examination for rotator cuff strength. An electromyographic investigation. *Am J Sports Med.* 1996;24(5):581-588.

85. Hayes K, Walton JR, Szomor ZL, Murrell GA. Reliability of 3 methods for assessing shoulder strength. *J Shoulder.Elbow.Surg.* 2002;11(1):33-39.
86. Rabin A, Irrgang JJ, Fitzgerald GK, Eubanks A. The intertester reliability of the Scapular Assistance Test. *J Orthop.Sports Phys Ther.* 2006;36(9):653-660.
87. Kibler WB. The role of the scapula in athletic shoulder function. *Am J Sports Med.* 1998;26(2):325-337.
88. Kibler WB, Uhl TL, Maddux JW, Brooks PV, Zeller B, McMullen J. Qualitative clinical evaluation of scapular dysfunction: a reliability study. *J Shoulder.Elbow.Surg.* 2002;11(6):550-556.
89. Borstad JD. Resting position variables at the shoulder: evidence to support a posture-impairment association. *Phys Ther.* 2006;86(4):549-557.
90. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407-415.
91. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56(5):395-407.
92. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Stat.Med.* 1998;17(14):1623-1634.
93. Brox JI, Gjengedal E, Uppheim G, et al. Arthroscopic surgery versus supervised exercises in patients with rotator cuff disease (stage II impingement syndrome): a prospective, randomized, controlled study in 125 patients with a 2 1/2-year follow-up. *J Shoulder.Elbow.Surg.* 1999;8(2):102-111.
94. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol.* Apr 1 2006;163(7):670-675.
95. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem.* 1993;39(4):561-577.
96. Hosmer D, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York: John Wiley and Sons, Inc; 2000.
97. Gulotta LV, Nho SJ, Dodson CC, Adler RS, Altchek DW, MacGillivray JD. Prospective evaluation of arthroscopic rotator cuff repairs at 5 years: part II--prognostic factors for clinical and radiographic outcomes. *J Shoulder Elbow Surg.* Sep;20(6):941-946.
98. Fritz JM, George SZ, Delitto A. The role of fear-avoidance beliefs in acute low back pain: relationships with current and future disability and work status. *Pain.* Oct 2001;94(1):7-15.
99. Hellum C, Johnsen LG, Gjertsen O, et al. Predictors of outcome after surgery with disc prosthesis and rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up. *Eur Spine J.* January 2012.
100. Lethem J, Slade PD, Troup JD, Bentley G. Outline of a Fear-Avoidance Model of exaggerated pain perception--I. *Behav Res Ther.* 1983;21(4):401-408.
101. Slade PD, Troup JD, Lethem J, Bentley G. The Fear-Avoidance Model of exaggerated pain perception--II. *Behav Res Ther.* 1983;21(4):409-416.
102. Kuijpers T, van der Windt DA, Boeke AJ, et al. Clinical prediction rules for the prognosis of shoulder pain in general practice. *Pain.* 2006;120(3):276-285.
103. Ryall C, Coggon D, Peveler R, Poole J, Palmer KT. A prospective cohort study of arm pain in primary care and physiotherapy--prognostic determinants. *Rheumatology.(Oxford).* 2007;46(3):508-515.

104. Mintken PE, Cleland JA, Carpenter KJ, Bieniek ML, Keirns M, Whitman JM. Some factors predict successful short-term outcomes in individuals with shoulder pain receiving cervicothoracic manipulation: a single-arm trial. *Phys Ther.* Jan;90(1):26-42.

105. George SZ, Fritz JM, Bialosky JE, Donald DA. The effect of a fear-avoidance-based physical therapy intervention for patients with acute low back pain: results of a randomized clinical trial. *Spine (Phila Pa 1976).* Dec 1 2003;28(23):2551-2560.

106. Lindstrom I, Ohlund C, Eek C, et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. *Phys Ther.* Apr 1992;72(4):279-290; discussion 291-273.

107. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain.* 2000;85(3):317-332.

108. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry.* Oct 2002;59(10):877-883.

109. Lin JJ, Lim HK, Yang JL. Effect of shoulder tightness on glenohumeral translation, scapular kinematics, and scapulohumeral rhythm in subjects with stiff shoulders. *J Orthop.Res.* 2006;24(5):1044-1051.

110. Muraki T, Aoki M, Uchiyama E, Murakami G, Miyamoto S. The effect of arm position on stretching of the supraspinatus, infraspinatus, and posterior portion of deltoid muscles: a cadaveric study. *Clin Biomech.(Bristol., Avon.).* 2006;21(5):474-480.

111. van der Windt DA, Kuijpers T, Jellema P, Van Der Heijden GJ, Bouter LM. Do psychological factors predict outcome in both low-back pain and shoulder pain? *Ann Rheum.Dis.* 2007;66(3):313-319.

112. Michener LA, McClure PW, Karduna AR. Anatomical and biomechanical mechanisms of subacromial impingement syndrome. *Clin Biomech.(Bristol., Avon.).* 2003;18(5):369-379.

113. Razmjou H, Holtby R, Myhr T. Gender differences in quality of life and extent of rotator cuff pathology. *Arthroscopy.* 2006;22(1):57-62.

114. Steyerberg E. *Clinical Prediction Models: a practical approach to development, validation, and updating*. New York: Springer; 2009.

115. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med.* May-Jun 2004;66(3):411-421.

116. Ring D, Kadzielski J, Fabian L, Zurakowski D, Malhotra LR, Jupiter JB. Self-reported upper extremity health status correlates with depression. *J Bone Joint Surg Am.* Sep 2006;88(9):1983-1988.

117. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine (Phila Pa 1976).* Dec 15 2000;25(24):3192-3199.

118. Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC.Musculoskelet.Disord.* 2003;4:11.

119. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol.* May 2000;53(5):459-468.

120. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care.* May 1999;37(5):469-478.

121. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* Sep 1999;52(9):861-873.
122. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis.* 1986;39(11):897-906.
123. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* Mar 1989;27(3 Suppl):S178-189.
124. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther.* Oct 1996;76(10):1109-1123.
125. Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol.* Mar 1997;50(3):239-246.
126. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum.* May 1985;28(5):542-547.
127. Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res.* Jun 1999;8(2):113-134.
128. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* Jan 2007;60(1):34-42.
129. Portney L, Watkins M. *Foundations of Clinical Research Applications to Practice*. Norwalk, Connecticut: Appleton & Lange; 1993.
130. Finch E, Brooks D, Stratford P, Mayo N. *Physical Rehabilitation Outcome Measures: a guide to enhanced decision making*. 2nd ed. Hamilton, Ontario: BC Decker; 2002.
131. Bellamy N, Carr A, Dougados M, Shea B, Wells G. Towards a definition of "difference" in osteoarthritis. *J Rheumatol.* Feb 2001;28(2):427-430.
132. Ekeberg OM, Bautz-Holter E, Keller A, Tveita EK, Juel NG, Brox JI. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. *J Clin Epidemiol.* May;63(5):575-584.
133. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol.* Oct 2004;57(10):1008-1018.
134. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract.* Feb 2000;6(1):39-49.
135. Beaton DE, van Eerd D, Smith P, et al. Minimal change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. *J Clin Epidemiol.* May 2011;64(5):487-496.
136. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis.* Jan 2005;64(1):34-37.
137. Christie A, Dagfinrud H, Garratt AM, Ringen Osnes H, Hagen KB. Identification of shoulder-specific patient acceptable symptom state in patients with rheumatic diseases undergoing shoulder surgery. *J Hand Ther.* Jan-Mar;24(1):53-60; quiz 61.
138. Khanna PP, Maranian P, Gregory J, Khanna D. The minimally important difference and patient acceptable symptom state for the Raynaud's condition score in patients with Raynaud's phenomenon in a large randomised controlled clinical trial. *Ann Rheum Dis.* Mar;69(3):588-591.

139. Razmjou H, Bean A, van OV, MacDermid JC, Holtby R. Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. *BMC.Musculoskelet.Disord.* 2006;7:26.