Masthead Logo

Nova Southeastern University
NSUWorks

Department of Physical Therapy Student Theses, Dissertations and Capstones

Department of Physical Therapy

1-1-2018

Shoulder Symptom Irritability: Development and Testing of a New Construct

Stephen Michael Kareha
Nova Southeastern University

This document is a product of extensive research conducted at the Nova Southeastern University College of Health Care Sciences. For more information on research and degree programs at the NSU College of Health Care Sciences, please click here.

Follow this and additional works at: https://nsuworks.nova.edu/hpd_pt_stuetd

Part of the Physical Therapy Commons

All rights reserved. This publication is intended for use solely by faculty, students, and staff of Nova Southeastern University. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, now known or later developed, including but not limited to photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the author or the publisher.

NSUWorks Citation

Stephen Michael Kareha. 2018. *Shoulder Symptom Irritability: Development and Testing of a New Construct.* Doctoral dissertation. Nova Southeastern University. Retrieved from NSUWorks, College of Health Care Sciences - Physical Therapy Department. (73) https://nsuworks.nova.edu/hpd_pt_stuetd/73.

This Dissertation is brought to you by the Department of Physical Therapy at NSUWorks. It has been accepted for inclusion in Department of Physical Therapy Student Theses, Dissertations and Capstones by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

المنارة للاستشارات

www.manaraa.com

Shoulder Symptom Irritability:
Development and Testing of a New Construct

by

Stephen Michael Kareha

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Nova Southeastern University
College of Health Care Sciences
Physical Therapy Department

2018

www.manaraa.com

-- SIGNATURE PAGE WILL BE INSERTED HERE --

www.manaraa.com

DISSERTATION ABSTRACT

Background: Physical therapists regularly make decisions regarding intervention intensity based upon pathoanatomy and symptom irritability, but the reliability and validity of classifying patients by symptom irritability are unknown.
Purpose: Examine the reliability and construct validity of the shoulder symptom irritability classification (SSIC) system for the purposes of determining an appropriate treatment intensity.
Design: Prospective repeated-measures cross-sectional single-blinded design.
Methods: 101 consecutive subjects with primary complaints of shoulder pain were assessed by a pair of blinded raters. Raters recorded the SSIC level and selected the appropriate intervention intensities for the subjects.
Data Analysis: Prevalence-adjusted, bias-adjusted Kappa for ordinal scales (PABAK-OS) and observed agreement were the primary measures of reliability. Analysis of variance (ANOVA) was used to compare functional disability across different levels of irritability. Receiver operating characteristic (ROC) curve analysis was utilized to derive cut-off scores for the patient-reported outcome (PRO) measures. Ordinal regression was utilized to compare the strength of patient-reported pain and disability in the determination of shoulder symptom irritability.
Results: Inter-rater reliability (PABAK-OS) was 0.69 (95% Confidence Interval [CI] = 0.59, 0.78). ANOVA demonstrated significant differences in functional limitation between SSIC groups for all PRO measures. ROC curve analysis found significant cut-off scores for all PRO measures. Lastly, rater agreement between SSIC and treatment strategy was found to have PABAK-OS of 0.82 (95% CI 0.75, 0.88) with 80% agreement.
Discussion: The inter-rater reliability of the SSIC system good and is not contingent upon experience or expertise. Despite lack of predominance of the function in the components of SSIC, functional limitation significantly influences SSIC along with aspects of pain that influence function. While the cut-off scores show promising results, further work is needed to validate the results. Ultimately, there appears to an excellent relationship between rater selected SSIC and treatment strategy demonstrating a foundation for construct validity of the SSIC. Therefore, the results of this study should serve as a foundation for future work for refinement of the SSIC as a component of the STAR-Shoulder diagnostic classification system.
Clinical Significance: The shoulder symptom irritability classification scale is reliable and clinically useful for improvement of communication between medical providers.

www.manaraa.com

ACKNOWLEDGEMENTS

It is with sincere gratitude that I have had the opportunity to work with and be supported by so many wonderful people. I want to thank my dissertation committee chair, Dr. Alicia Fernandez-Fernandez, for her tireless advisement, support, and encouragement throughout the dissertation process. I appreciate all your efforts in keeping me on track and facilitating me through this wonderful labyrinth of processes that is the dissertation. Thank you to my content expert, Dr. Philip McClure, for his sage advice and insight into understanding shoulder pain and the intricacies of novel classification systems. Thank you to my statistics expert, Dr. Samuel Cheng, for your wisdom in helping me through the understanding of statistical analyses. I have learned so much from all of you.

I also want to thank the many physical therapists throughout the St. Luke's University Health Network who worked to collect data for this project, I could not have done it without you. And most importantly, I want to thank my wife, Jennifer, my children, Savannah, Jocelyn, and Juliette, and my family for all your love and support over these past years.

It is said that we build upon the shoulders of giants and throughout this Ph.D. process, I have realized how true that statement really is. Soli Deo gloria!

www.manaraa.com

Table of Contents

DISSERTATION ABSTRACT iii
ACKNOWLEDGEMENTS iv
LIST OF FIGURES vi
LIST OF TABLES viii
CHAPTER 1: INTRODUCTION 1
Introduction 1
Statement of the problem investigated and the goal achieved 1
Relevance, Significance or Need for the Study 6
Elements, Hypotheses, Theories, or Research Questions Investigated 8
Definitions of Terms 9
Summary 18
CHAPTER 2: REVIEW OF THE LITERATURE 19
Introduction 19
Historical Overview of the Theory and Research Literature 19
The Theory and Research Literature Specific to the Topic 23
Summary of What Is Known and Unknown About the Topic 26
The Contribution This Study Makes to the Field 27
Summary 27
CHAPTER 3: METHODOLOGY 29
Introduction 29
Research Methods Employed 29
Specific Procedures Employed 31
Formats for Presenting Results 42
Resources Used 42
CHAPTER 4: RESULTS 44
Introduction 44
Findings 44
Summary of Results 74
CHAPTER 5: DISCUSSION 75
Introduction 75
Discussion 75

www.manaraa.com

Implications....92
Recommendations....94
Limitations and Delimitations....97
Summary....100
APPENDICES....109
Appendix A: Recruitment & Logistics....109
Appendix B: Rater Information....110
Appendix C: Inclusion/Exclusion Criteria....111
Appendix D: FOTO FS....112
Appendix E: PSS/ASES....115
Appendix F: Examination....117
Appendix G: Shoulder Symptom Irritability Classification....118
Appendix H: Treatment Strategy....119
Appendix I: Intervention Intensity....120
Appendix J: Inter-rater Contingency Tables for Each Site....121
Appendix K: Contingency Tables for Shoulder Symptom Irritability and Treatment Strategy....124
REFERENCES....132

www.manaraa.com

LIST OF FIGURES

Figure 1 Specific aims of the study……..……..….8

Figure 2 Participant inclusion in the analysis for this study..…….……...….44

Figure 3 Rater Age Distribution ...…….……...…....46

Figure 4 Rater Years of Practice Distribution …….…….…….....46

Figure 5 Inter-Rater Reliability Compared by Site……...…….....49

Figure 6 Inter-rater reliability controlling for specialist-trained raters……..... 51

Figure 7 Functional limitation differences between shoulder symptom irritability groups…….…….....52

Figure 8 Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for PSS Total Score…….…….....54

Figure 9 Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for PSS Total Score.……...….....55

Figure 10 Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for ASES Total Score..….……55

Figure 11 Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for ASES Total Score...……..…….....56

Figure 12 Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for FOTO FS…..……..….56

Figure 13 Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for FOTO FS…….…...57

Figure 14 Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for PSS Function Subscale..……....57

Figure 15 Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for PSS Function Subscale..…….…..……..58

www.manaraa.com

Figure 16 Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for ASES Function Subscale……....58

Figure 17 Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for ASES Function Subscale ……...59

Figure 18 Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for PSS Pain Subscale …….....59

Figure 19 Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for PSS Pain Subscale ……....….....60

Figure 20 Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for ASES Pain Subscale …….....60

Figure 21 Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for ASES Pain Subscale.……61

Figure 22 Distribution of PSS total score based upon rater selection.……....62

Figure 23 Distribution of ASES total score based upon rater selection…..…….....63

Figure 24 Distribution of FOTO Functional Score based upon rater selection…...63

Figure 25 Correlation between PSS Pain Subscale and Shoulder Symptom Irritability.……..….…….....66

Figure 26 Correlation between PSS Function Subscale and Shoulder Symptom Irritability…….....67

Figure 27 Rater Agreement between selected Shoulder Symptom Irritability and Treatment Strategy…….....91

www.manaraa.com

LIST OF TABLES

Table 1 Intervention choices based upon treatment intensity……..17

Table 2 Proposed factors of symptom irritability……..……..... 25

Table 3 Rater demographics……..……..... 45

Table 4 Patient demographics ……..……....47

Table 5 Frequency of Shoulder Symptom Irritability Rating..... 48

Table 6 Inter-Rater Reliability of Shoulder Symptom Irritability Classification – All Sites.……...…….....48

Table 7 Inter-Rater Reliability of Shoulder Symptom Irritability Classification by Site……... 49

Table 8 Inter-Rater Reliability of Shoulder Symptom Irritability Classification - Sites with n < 10……..……...…….... 50

Table 9 Inter-Rater Reliability of Shoulder Symptom Irritability Classification - Sites with n ≥ 10……..…….....50

Table 10 Functional limitation differences between shoulder symptom irritability groups.……...……...…….. 52

Table 11 Effect size of functional limitation between shoulder symptom irritability groups……..…….....53

Table 12 Receiver operating characteristic curve results ……...…….. 61

Table 13 Patient-reported outcome measure cut-off scores…….……...…..……..62

Table 14 Parameter Estimates for Ordinal Regression of PSS Pain Subscale and ASES Pain Subscale ……..…..……. 64

Table 15 Parameter Estimates for Ordinal Regression of PSS Pain Subscale items.....65

www.manaraa.com

Table 16 Parameter Estimates for Ordinal Regression of Pain and Function..... 66

Table 17 Agreement between selected Shoulder Symptom Irritability and Treatment Strategy..... 68

Table 18 Agreement of Specialists between selected Shoulder Symptom Irritability and Treatment Strategy ……..……....68

Table 19 Agreement of Non-Specialists between selected Shoulder Symptom Irritability and Treatment Strategy..... 69

Table 20 Agreement of Raters with more than 5 years of experience between selected Shoulder Symptom Irritability and Treatment Strategy……..……...….. 70

Table 21 Agreement of Raters with 5 or fewer years of experience between selected Shoulder Symptom Irritability and Treatment Strategy..... 70

Table 22 Agreement of Raters with more than 10 years of experience between selected Shoulder Symptom Irritability and Treatment Strategy..... 71

Table 23 Agreement of Raters with 10 or fewer years of experience between selected Shoulder Symptom Irritability and Treatment Strategy……..……. 71

Table 24 Agreement between selected Shoulder Symptom Irritability and Treatment Strategy…..... 72

Table 25 Agreement between selected Shoulder Symptom Irritability and Treatment Strategy – Outliers removed……..……..……..... 73

Table 26 Agreement between selected Shoulder Symptom Irritability and Treatment Strategy – Largest site removed…….……..…….. 73

www.manaraa.com

CHAPTER 1: INTRODUCTION

Introduction

Shoulder disorders are a large medical and economic problem throughout the world that results in pain, functional limitations, and disability.[1,2] Unfortunately, shoulder disorders are frequently recurrent, resulting in greater than 50% of patients continuing to have pain and limitations more than a year after onset.[3,4]

Statement of the problem investigated and the goal achieved

Clinical Decision-Making

Clinicians make decisions regarding the intensity of interventions based upon diagnosis. Traditionally, the diagnosis is based on pathoanatomy and thus is assumed to differentiate patients into homogenous groups to guide treatment and prognosis.[2] While this pathoanatomic system of diagnosis may be adequate for surgical decision-making, due to the anatomic restoration achieved with surgical procedures, it may be inadequate for non-surgical decision making.[2,5] Additionally, recent evidence has demonstrated a poor correlation between pathoanatomic diagnosis and the selected non-surgical interventions by orthopedic clinical specialists, in terms of their ability to effectively resolve patients' functional limitations and disability.[6]

Due to this lack of correlation between pathoanatomic diagnosis and selection of non-surgical interventions, the Staged Approach for Rehabilitation Classification: Shoulder Disorders (STAR-Shoulder) classification scheme has been proposed to enhance clinical decision-making.[2] The STAR-Shoulder utilizes pathoanatomic diagnostic classification paired with identified physical impairments and symptom

www.manaraa.com

irritability to more appropriately direct treatment decision-making for shoulder disorders.[2]

Others have also suggested[7-9] that symptom irritability should be assessed to appropriately dose the stress to the tissues of the body, as symptom irritability may be an indicator of the degree of inflammation. Thus, in conjunction with pathoanatomic classification and type of impairments, symptom irritability would provide a more consistent framework from which a clinician could make clinical decisions.[2] However, without a reliable classification system for determining shoulder symptom irritability, the clinical decision-making for determining the intensity of physical stress to tissue is much more challenging and inconsistent.

Physical Stress Theory

The Physical Stress Theory (PST)[10] describes changes in the ability for tissues to adapt to changes in stress based upon movement and alignment factors, extrinsic factors, psychosocial factors, and physiological factors, including inflammation. The PST suggests that biologic tissue will remodel according to stresses applied to them.[10]

The PST also postulates that inflammation and injury lower the threshold for tissue adaptation, and consistent overload of the tissues elevates the threshold for tissue adaptation.[10] Therefore, after biologic tissues are injured, simple premorbid activities may induce injurious stresses to the tissues. Symptom irritability is a construct that reflects this ability, or inability, of tissues to handle physical stress.[2,7,8] It is important that recently injured and inflamed tissues are protected from subsequent excessive stress until acute inflammation resolves.

www.manaraa.com

There are currently no clinical markers for the level of inflammation available, and measurements such as erythrocyte sedimentation rate, C-reactive protein, and plasma viscosity testing are impractical in clinical situations. In practical scenarios, the construct of symptom irritability is utilized by clinicians to determine the intensity of examination and intervention.[11,12] Therefore, it is imperative that the construct of symptom irritability be developed to reliably measure the thresholds for appropriate tissue adaptations,[10] as this can help to avoid further injury and increase the effectiveness of clinical intervention.

Construct of Symptom Irritability

Multiple experts in physical therapy have proposed criteria for symptom irritability from which to base clinical decisions for intensity.[7-9,13,14] Maitland described the measurement of irritability via the relationship of (1) the vigor of activity required to provoke a patient's symptoms, (2) the severity of those symptoms, and (3) the time it takes for the symptoms to subside once aggravated (i.e., pain persistence).[13,15]

The reliability of Maitland's classification was recently tested and found to be poor to moderate.[13,16] The construct of symptom irritability was further studied by Barakatt and colleagues[13,15] based upon the ranking of pre-defined factors. However, the resulting classification scheme including disability/pain intensity, pain persistence, sitting limit, standing limit, forward bend limit, and walking/lifting limit has not been validated or studied for its reliability.[13,15] As this construct was analyzed with regard to low back pain, many of the factors that Barakatt and colleagues[13,15] proposed are specific to low back pain and may not be the same factors for symptom irritability of shoulder pain. Additionally, this study did not acquire the factors from a large sample with varying backgrounds, resulting in significant limitations to the external validity of the study.

www.manaraa.com

The symptom irritability construct was also described in the decision-making process for intervention selection for low back pain by Delitto and colleagues via the relationship of (1) time since injury, (2) level of disability, and (3) psychological sequelae.[8] While this classification was not directly described as irritability, but rather described it as acuity, it constitutes similar characteristics and has been a cornerstone to the utility of the Treatment-Based Classification system. This concept has been perpetuated by the clinical practice guideline for low back pain[17] and multiple papers describing the treatment based classification system[18-26] that developed from the 1995 paper.

Finally, Kelley and McClure proposed a method of classifying symptom irritability specifically for the shoulder. This method of symptom irritability has been supported by the Shoulder Pain and Mobility Deficits: Adhesive Capsulitis: Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability, and Health From the Orthopaedic Section of the American Physical Therapy Association.[27] Additionally, in an effort to improve clinical decision-making for non-operative management the proposed Staged Approach for Rehabilitation Classification: Shoulder Disorders (STAR-Shoulder) system utilized the patient's pathoanatomic diagnosis, shoulder symptom irritability level, and physical impairments to determine the most effective treatment.[2]

Knowledge Gap

While symptom irritability has been described by clinicians[9,28] and researchers,[8,14,16] to my knowledge, no studies have determined the reliability and validity of symptom irritability measurement. This knowledge gap is surprising, due to emerging

www.manaraa.com

evidence that adherence to guidelines incorporating classification based upon acuity, or symptom irritability, for other body regions significantly reduces health care utilization and cost,[29] and the potential importance of symptom irritability in guiding intervention for shoulder disorders.[2]

Long-Term Goal

The long-term goal of my research agenda is to determine the reliability and construct validity of the Staged Approach for Rehabilitation Classification: Shoulder Disorders (STAR-Shoulder)[2] and refine the system for non-surgical clinical decision-making for shoulder disorders and other body regions.

Purpose of Dissertation

The objective of this dissertation is to begin to establish the reliability and construct validity of shoulder symptom irritability as one part of the STAR-Shoulder classification system to guide refinements. The central hypothesis is that shoulder symptom irritability is a reliable classification system that directs treatment intensity. This hypothesis has been formulated on the basis of the Physical Stress Theory[10] and studies on spinal pain[13,15,16] that symptom irritability is a marker of tissue readiness for physical stress. This central hypothesis has been framed further by studies on spinal pain[30] and expert consensus[8,14,27] which have purported the use of classification systems incorporating symptom irritability improve patient outcomes. However, the proposed classification system for the shoulder is only at the conceptual stage, and thus research is required to refine and validate the proposed models.[2]

The rationale for this research is that when shoulder symptom irritability is appropriately measured and communicated, non-surgical interventions can be prescribed

www.manaraa.com

at the appropriate intensity. Thus, the classification scheme must first be tested for reliability between clinicians for greater generalizability. Subsequently, the shoulder symptom irritability classification system must be evaluated for validity. Finally, the relationship between self-reported functional limitations and therapist judgments of irritability needs to be further clarified for improved comparison between shoulder symptom irritability groups.

Relevance, Significance or Need for the Study

The validity of this shoulder symptom irritability classification system was questioned by a study investigating intervention prescription for adhesive capsulitis as outcomes were no differences among groups of differing levels of symptom irritability for the same intervention intensity.[31] However, Dempsey and colleagues[31] utilized a retrospective post-operative cohort sample (n=36) from a single orthopedic surgeon, which calls into question the generalizability of their findings. This is problematic as there are significantly different aspects of post-operative care compared to non-operative care, including operative technique and time since surgery, which may supersede symptom irritability. Furthermore, while Dempsey and colleagues[31] utilized the same basic list of criteria for shoulder symptom irritability as Kelley and McClure[14], Dempsey and colleagues excluded subjects from the low irritability classification even if they had four criteria specifying low irritability and only one criterion indicating higher irritability levels.[31] Thus, the internal validity is called into question as the classification scheme described by Dempsey et al[31] would allow subjects to be classified as moderate/high irritability simply due to a single aberrant characteristic, instead of the intended use of the cluster of criteria to develop an overall classification of shoulder symptom irritability.

www.manaraa.com

On the contrary, other evidence supports this notion that intensity should be altered based upon symptom irritability. A prospective cohort study in 2004 (n=77) demonstrated significantly worse self-reported functional outcomes with a protocol ignoring symptom irritability than with a program accounting for symptom irritability.[32] This is further supported by biologically based reasoning that intensity should be altered based upon symptom irritability.

Current evidence for the definition of shoulder symptom irritability includes expert consensus and clinical commentary. This project is innovative because it establishes the reliability of shoulder symptom irritability classification and begins to establish a correlation between shoulder symptom irritability and guidance of treatment decision making. Furthermore, it concurrently provides an efficient method of communication between providers of all healthcare professions to better manage non-operative patient care.

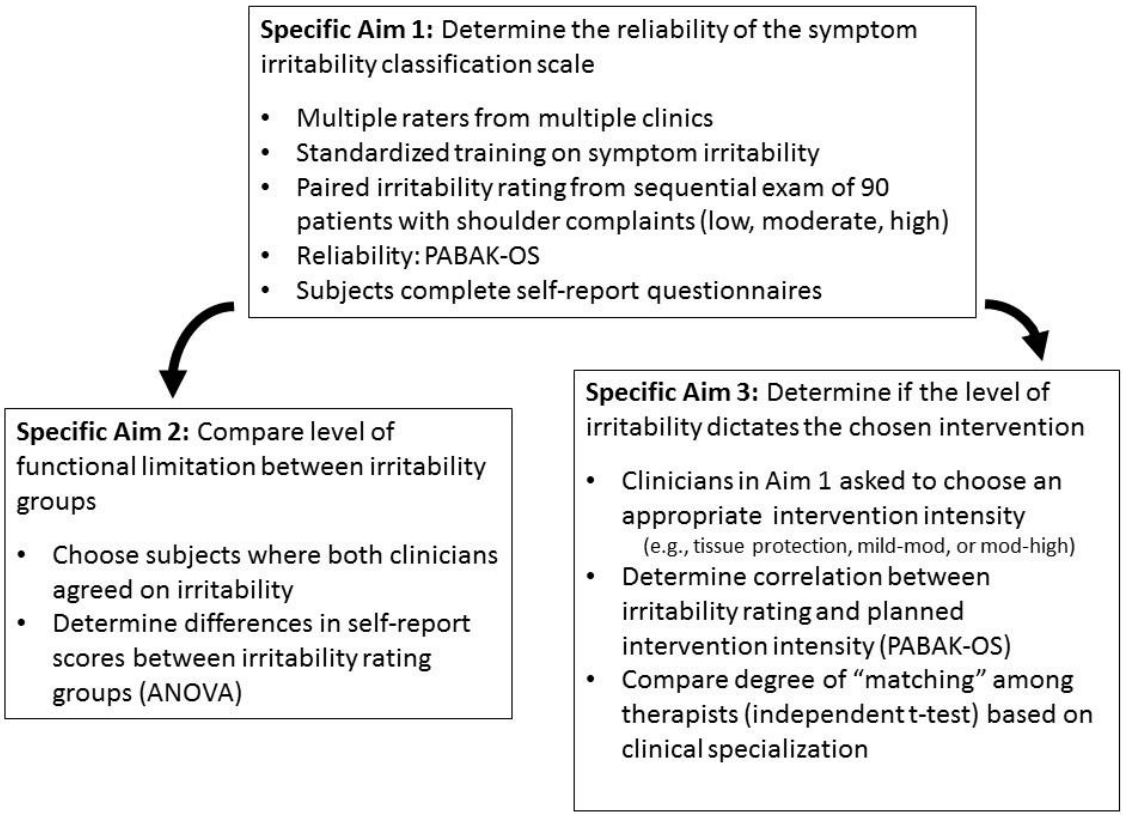
This approach could help us move beyond the current state of heterogeneous diagnostic groups, and improve the effectiveness of intervention to aid in managing rising healthcare costs. A reliable and valid shoulder symptom irritability classification system needs to be integrated with the International Classification of Functioning, Disability and Health (ICF) categories of health condition (pathoanatomy) and body functions and structure (impairments) to appropriately prescribe rehabilitation intervention and reduce unwarranted variation in clinical practice. Ultimately, the reduction in unwarranted variation has the potential to result in reduced costs and improved functional outcomes for patients.

www.manaraa.com

Elements, Hypotheses, Theories, or Research Questions Investigated

The study tested the central hypothesis that shoulder symptom irritability, a component of the STAR-Shoulder classification scheme,[2] is a reliable classification system that directs treatment intensity with the following three specific aims (Figure 1).

Figure 1: Specific Aims of the Study



1. Specific Aim 1: The first specific aim was to determine the reliability of the shoulder symptom irritability classification system. The hypothesis for Specific Aim 1 was that the shoulder symptom irritability classification system demonstrates good reliability (K>0.60 and agreement>70%) between raters.
2. Specific Aim 2: The second specific aim was to compare levels of functional limitation between shoulder symptom irritability groups. The two hypotheses

www.manaraa.com

for Specific Aim 2 were: 1) there is a significant difference ($p<0.05$) in patient-reported functional limitations between shoulder symptom irritability groups; and 2) pain subscales demonstrate stronger differences than functional subscales between shoulder symptom irritability groups.

3. Specific Aim 3: The third specific aim was to determine if the level of shoulder symptom irritability dictates the chosen intervention intensity. The two hypotheses for Specific Aim 3 were: 1) the level of shoulder symptom irritability is moderately correlated ($K>0.40$ and agreement>50%) with planned intervention intensity; and 2) clinicians with clinical specialization (e.g. OCS, FAAOMPT) have a significantly higher degree ($p<0.05$) of matched planned intervention intensity compared to those without a specialization.

All outcome measures and intervention choices used in the study are described in the upcoming definitions section, and in more detail in subsequent chapters.

Definitions of Terms

Computerized Adaptive Testing

Computerized adaptive testing is a method of administering examinations to increase the efficiency of the examination process by re-estimating the testee's ability level each time an answer is selected. This is done utilizing item response theory to evaluate each item and item response, such that the estimate of the testee's ability or disability level becomes more precise each time a response is provided.[33] This study utilized computerized adaptive testing for the construct of fear avoidance, as described later on in the chapter.

www.manaraa.com

Degree of Disability

The degree of disability was measured utilizing the Penn Shoulder Scale, the American Shoulder and Elbow Surgeons Disability Scale, and the Focus on Therapeutic Outcomes (FOTO) Functional Score, which are patient-reported disability measures. As no patient-reported disability measures have been deemed the gold standard for persons with shoulder pain,[34,35] this study included multiple measures. Additionally, neither of these scales have well-defined ranges for low, moderate, and high disability,[2,34] and thus the clinicians were asked to use their judgment to determine the meaning of the score for purposes of classifying a patient's symptom irritability.

Dry Needling

Dry needling uses a thin filiform needle without medication to penetrate the skin and stimulate underlying myofascial trigger points, contractile tissues, and connective tissues for the management of neuromusculoskeletal pain and movement impairments.[36]

Efficiency

The efficiency of a clinician's performance was defined by the mean number of visits utilized for an episode of care marked as "shoulder" and in the Focus on Therapeutic Outcomes (FOTO) database.

Electrical Agents

Electrical agents include interventions such as laser, pulsed electromagnetic field, and electrical modalities aimed at modulating pain or eliciting a muscular contraction.[37-39]

End Feel

End feel is generally defined as the sensation perceived by the clinician when resistance to motion is felt and whether that sensation of resistance is due to pain or tissue

www.manaraa.com

tension.[40] In this study, end feel was specifically be used to indicate the onset of pain in relation to onset of tissue resistance. While the ability to utilize end feel to determine sequence of pain in relation to tissue resistance has demonstrated variable inter-rater reliability (K = 0.62 to 0.76,[41] K_w = -0.01 to 0.70[40]), it has generally shown good intra-rater reliability (K = 0.48 to 0.59,[41] K_w = 0.59 to 0.87[40]) and is frequently used for clinical decision-making.[40]

Examination

The examination encompassed any tests and measures required to determine shoulder symptom irritability, but was to avoid any symptom altering procedures prior to both examiners completing their assessments.

Frequency

Frequency is related to how often an intervention is performed, measured in number of sessions per week.

Instrument-Assisted Soft Tissue Mobilization

Instrument-assisted soft tissue mobilization is defined as a manual therapy technique performed with ergonomically designed instruments, comprising a continuum of skilled passive movements to the soft tissue that is applied at varying speeds and amplitudes.

Manual Soft Tissue Mobilization

Manual soft tissue mobilization is defined as a manual therapy technique comprising a continuum of skilled passive movements directed at muscular and connective tissue that are applied at varying speeds and amplitudes. Examples include, but are not limited to, deep pressure and various massage techniques.

www.manaraa.com

Neuromuscular Control/Coordination Training

Neuromuscular control/coordination training is defined as procedures or exercises designed to retrain the movement pattern[42] of the shoulder girdle, spine, and/or other interdependent body regions. This training focuses on precision and quality of movement rather than overload. At this time, the literature shows strong evidence for the use of neuromuscular control and coordination exercises.[37]

Patient Education/Activity Modification

Patient education, counseling, and activity modification can be done in a variety of ways. Media such as pamphlets, videos, and verbal advice have been assessed in the current literature. Additionally, demonstrations with and without verbal and/or tactile cueing are frequently utilized in clinical practice. At this time, the literature shows moderate evidence for the use of patient education and counseling for patients who have suffered from adhesive capsulitis[27] and emerging evidence for patients with rotator cuff syndrome.[43,44]

Psychological Sequelae

Psychological sequelae include constructs such as fear avoidance, self-efficacy, catastrophization, and kinesiophobia. In this study, fear avoidance was the indicator of psychological sequelae, measured utilizing a computerized adaptive testing version of the Fear-Avoidance Beliefs Questionnaire. This questionnaire has demonstrated good ability to dichotomize people into high and low levels of fear avoidance.[45]

Resistive Strength Training Exercises (including isometric)

Resistive strength training exercises are defined as interventions that intend to increase strength and/or endurance of muscles including isometric, isotonic, and

www.manaraa.com

isokinetic movements. Strengthening exercises may begin in a protected mid-range position with the limb supported, and progress to end-range positions that work against gravity and additional external resistance. Exercise is progressed based on variables such as repetitions, resistance, speed and complexity of movement, body and joint position, and timing of muscular activation.[42] Strength training specifically involves overloading the muscle and exercising until fatigue is achieved.[37]

Range of Motion (ROM) Exercises (end range)

ROM Exercises are defined as a continuum of therapeutic movements, either manually applied by the clinician or performed by the patient, directed at moving the shoulder girdle through the physiologic range of motion. The end range category includes all movements that aim at reaching end range of movement, but do not include those techniques aimed at maintaining end range positioning for longer periods of time.

Range of Motion (ROM) Exercises (non-end range)

ROM Exercises are defined as a continuum of therapeutic movements, either manually applied by the clinician or performed by the patient, directed at moving the shoulder girdle through the physiologic range of motion. The non-end range category includes all movements that avoid end range of movement, usually prescribed to facilitate pain reduction and fluidity of joint movement while avoiding end-range stress on tissue.

Range of Motion (ROM)/Stretching Exercises (long duration)

ROM Exercises are defined as a continuum of therapeutic movements, either manually applied by the clinician or performed by the patient, directed at moving the shoulder girdle through the physiologic range of motion. The stretching exercises category includes all movements that aim at providing end-range stress to increase

www.manaraa.com

movement and utilize end range positioning for longer periods of time, typically between 30 seconds and several minutes.

Shoulder: Joint Mobilization – End Range

Shoulder: Joint mobilization – End range is defined as a manual therapy technique directed at the shoulder girdle, comprising a continuum of skilled passive movements to the joints that are applied at varying speeds and amplitudes, including a small amplitude/high-velocity therapeutic movement, and are aimed at encountering tissue resistance.[27,46]

Shoulder: Joint Mobilization – Non-End Range

Shoulder: Joint mobilization – Non-end range is defined as a manual therapy technique directed at the shoulder girdle, comprising a continuum of skilled passive movements to the joints that are applied at varying speeds and amplitudes but NOT encountering tissue resistance.[27,46]

Spinal Manipulation (thrust)

Spinal manipulation (Thrust) is defined as a manual therapy technique directed at the cervical, thoracic, or lumbar spine, comprising a continuum of skilled passive movements to the joints that are applied utilizing a small amplitude/high-velocity therapeutic movement.[27,46]

Spinal Mobilization (non-thrust)

Spinal mobilization (Non-thrust) is defined as a manual therapy technique directed at the cervical, thoracic, or lumbar spine, comprising a continuum of skilled passive movements to the joints that are applied at varying speeds and amplitudes, excluding small amplitude/high-velocity therapeutic movements.[27,46]

www.manaraa.com

Shoulder Symptom Irritability

Symptom irritability is defined as tissue readiness to accept physical stress.[8,14,15] Providers were instructed to choose one of three levels of shoulder symptom irritability (high, moderate, or low) considering criteria including pain level, presence of night/resting pain, onset of pain during motion, comparison of active and passive mobility, and disability level.[2,27]

Taping/Strapping

Taping or strapping interventions include those techniques utilizing tape with varying levels of adhesiveness and elasticity to facilitate or inhibit specific joint movements, muscle function, and/or motor coordination.

Therapeutic Ultrasound

Therapeutic ultrasound is the use of sound waves to produce heating of deeper tissues (including muscles, tendons, ligaments, and scar tissue) and alteration of cellular activity (acoustical streaming and stable cavitation).[37]

Thermal Modalities

Thermal modalities included dry and moist hot pack application, ice and cold pack application, ice massage, and diathermy.

Treatment Intensity

Treatment intensity is defined as the amount of force necessary to perform the intervention. This is multifactorial in nature and thus depended on the specific intervention involved.

www.manaraa.com

Low-Intensity Interventions

Examples of low-intensity interventions would include activity modification and support to avoid further irritation, pain-free and non-end range mobility exercises and mobilizations and passive modalities. Specific interventions included in this category are listed in Table 1.

Moderate-Intensity Interventions

Examples of moderate-intensity interventions would include activity modification to progressively load the injured tissues without overload, comfortable end-range mobility exercises and mobilizations, movement training with emphasis on motor coordination/quality of motion, light to moderate resistance exercises to fatigue with avoidance of end range, and limited passive modality use. Specific interventions included in this category are listed in Table 1.

High-Intensity Interventions

Examples of high-intensity interventions would include no use of passive modalities, tolerable and longer duration and frequency of end range mobility exercises and mobilizations, high demand movement training with emphasis on motor coordination/quality of motion, and moderate to high resistance exercises to fatigue to include movements into end range.[2] Specific interventions included in this category are listed in Table 1.

www.manaraa.com

Table 1: Intervention Choices based upon Treatment Intensity

Low Intensity	Moderate Intensity	High Intensity
Shoulder: Joint Mobilization – Non-end range	Shoulder: Joint Mobilization – End range	Shoulder: Joint Mobilization – End range
Spinal Mobilization (Non-thrust)	Spinal Mobilization (Non-thrust)	Spinal Mobilization (Non-thrust)
Spinal Manipulation (Thrust)	Spinal Manipulation (Thrust)	Spinal Manipulation (Thrust)
Manual Soft Tissue Mobilization	Manual Soft Tissue Mobilization	Manual Soft Tissue Mobilization
	Instrument-Assisted Soft Tissue Mobilization	Instrument-Assisted Soft Tissue Mobilization
	Dry Needling	
	Neuromuscular Control/Coordination Training	Neuromuscular Control/Coordination Training
Range of Motion (ROM) Exercises (non-end range)	Range of Motion (ROM) Exercises (end range)	Range of Motion (ROM) Exercises (overpressure/long duration)
	Resistive Strength Training Exercises (including isometric)	Resistive Strength Training Exercises (including isometric)
Taping/Strapping	Taping/Strapping	
Patient Education/Activity Modification	Patient Education/Activity Modification	Patient Education/Activity Modification
Therapeutic Ultrasound		
Electrical Agents	Electrical Agents	
Thermal Modalities	Thermal Modalities	

Interventions in each column are considered matched to treatment intensity listed at the top of the column.

Vigor of Activity to Provoke Symptoms

The vigor of activity to provoke symptoms includes such measures as pain at rest, the degree of pain with activity, the presence of pain prior to end range movement, and tolerance to motion.

www.manaraa.com

Summary

As the healthcare system is struggling to determine the most cost-effective care for musculoskeletal conditions, it is imperative that clinical decision-making for non-operative shoulder disorders be improved. The STAR-Shoulder classification system has been proposed to improve clinical decision-making for non-operative shoulder disorders by utilizing a three-pronged approach: pathoanatomic classification, shoulder symptom irritability classification, and impairment classification.[2] While many clinicians and researchers have recommended the use of symptom irritability for determining the intensity of proposed interventions to alleviate shoulder pain, no studies have attempted to address the reliability and validity of classifying patients based upon shoulder symptom irritability.[7-9,13,14] This study determined the reliability of the shoulder symptom classification scale and begin to understand the correlation between shoulder symptom classification and treatment intensity decisions. Furthermore, this study also aided in the understanding the relationship between the degree of functional limitation and therapist judgment of shoulder symptom irritability.

www.manaraa.com

CHAPTER 2: REVIEW OF THE LITERATURE

Introduction

The theory of tissue irritability and symptom irritability is rooted in the stages of acute tissue healing. However, it was not until the late twentieth century that a physiotherapist from New Zealand began teaching others to utilize the concept of tissue irritability to gauge the intensity of treatment.[28] This symptom irritability is intended to be an indicator of the tissue's ability to handle physical stress.[2] In other words, it is a metric to clinically assess the degree of inflammatory activity present in order to guide appropriate intervention intensity.

Historical Overview of the Theory and Research Literature

Diagnosis has been integral in western medicine and is aimed at guiding the treatment approach, determining a prognosis, and succinctly communicating the signs and symptoms of the patient to other providers to aid in the patient's recovery.[47] Historically, diagnostic categories have been based solely upon pathoanatomy. To facilitate the accuracy of diagnosing pathoanatomy, much research has been performed to determine the reliability and validity of clinical testing[48-52] and imaging modalities.[49,52-55]

For a diagnosis to be meaningful, it is implicit that a diagnosis should direct the most appropriate intervention for that condition, determine a prognosis, and diagnoses should be mutually exclusive from one another. However, recent evidence suggests that pathoanatomic diagnosis is not correlated with the interventions chosen by board-certified specialists in orthopedic physical therapy.[6] Furthermore, pathoanatomic-based diagnosis may not be the best indicator for the determination of treatment strategy due to poor uniformity in labeling of shoulder disorders. Essentially, this poor uniformity of

www.manaraa.com

labeling creates heterogeneous groups of persons with shoulder pain instead of the intended homogenous groups.[47] This ambiguity results in an inability to effectively compare study results to determine the most effective treatments to maximize success rates.[47]

Even when pathology is classified appropriately, evidence has demonstrated a lack of correlation in activity limitations, participation restrictions, and symptoms. In a recent study investigating the correlation between clinical symptoms and power and function, there was poor correlation between degree of fatty degeneration of the rotator cuff muscles and power, as well as a poor correlation between degree of fatty degeneration and function.[56] Furthermore, even the degree of acromioclavicular joint osteoarthritis has been shown to have a poor correlation with clinical symptoms or even the side affected by clinical symptoms.[57]

Additionally, the diagnosis should direct the most appropriate intervention for a given condition. An example is in patients with chronic, symptomatic, full-thickness rotator cuff tear. There has been considerable debate in the literature regarding the most appropriate treatment for this very specific pathology, with some advocating for effective non-operative management[58] and others strongly recommending surgical repair.[59] It may seem logical to anticipate that the severity of rotator cuff tear pathology, such as the size of the tear and degree of retraction, would be highly predictive of the need for surgical repair in patients with chronic, symptomatic, full-thickness rotator cuff tear. However, no association was found between these pathoanatomic variables and the failure of non-operative rehabilitation in a recent study.[60]

www.manaraa.com

Therefore, a need exists to develop an adequate diagnostic system beyond the single classification construct of the anatomic structure implicated, in order to more accurately guide treatment decision making and inform prognosis.[2] Diagnostic classification systems designed to guide non-operative rehabilitation have been developed for the lumbar spine[8,61] and cervical spine.[62,63]

In 1995, a treatment based diagnostic process was proposed for the non-operative management of low back pain.[8] The proposed diagnostic system utilizes symptom acuity classification and physical impairment classification to determine the most accurate diagnostic classification to direct treatment decision making and is mutually exclusive.[8] From its original proposal, this system has been refined over the years[21,24,25,61,64] but still utilizes the same components of symptom acuity classification and physical impairment classification to determine diagnosis, prognosis, and most appropriate treatment intervention choice and intensity. In this system, symptom acuity is described as acute, subacute, and chronic, but despite the nomenclature utilized, it is notated that the acuity of symptoms is more related to the symptom irritability than the time since injury.[8] Furthermore, when comparing treatment matched to the diagnostic category to unmatched treatment, patients receiving matched treatment have demonstrated improved outcomes[18,20] and decreased healthcare costs[29] when utilized for acute and subacute low back pain.

Again, due to the heterogeneity of neck pain and resultant poor outcomes of interventional studies, a similar approach has been utilized for neck pain.[62,63,65] Diagnostic groups are separated based on symptom irritability and physical impairments.[62,63,65] In one diagnostic classification system, those patients with high

www.manaraa.com

symptom irritability were placed a separate category.[62] As this diagnostic classification system evolved though, the symptom irritability level became enmeshed with the physical impairment categories to direct treatment intensity.[63] When applied to patients with neck pain, this diagnostic classification system also produced superior outcomes when treatment was matched to the diagnostic category compared to unmatched treatment.[66]

Interestingly, due to the poor predictive value of specific pathology for lumbar and cervical spine disorders for the appropriate determination of non-operative management, neither the lumbar spine guidelines nor the cervical spine guidelines utilize pathoanatomical classification in the decision-making process for the most effective non-operative management.[61,63] The only utility of pathoanatomy in both lumbar and cervical spine diagnostic classification guidelines is during the screening process in order to determine appropriateness of the patient's condition for non-operative care and determine the need for referral to another health care provider.[18,20]

On the other hand, literature on the prognosis of shoulder disorders does demonstrate a correlation between pathoanatomic diagnostic classification and prognosis.[27,59,67-70] Thus, if the major aims of diagnosis are to direct treatment decisions and inform prognosis, it would not be prudent to ignore the implicated anatomical structures when diagnosing shoulder pain. Rather, the addition of shoulder symptom irritability classification and physical impairment classification to the pathoanatomic classification would provide a more complete diagnosis that both directs treatment and is mutually exclusive.[2]

www.manaraa.com

Thus, an optimal classification system to improve treatment decision-making would encompass pathoanatomy, shoulder symptom irritability, and physical impairments. The STAR-Shoulder classification system has been proposed to meet this need, utilizing pathoanatomic diagnostic classification paired with identified physical impairments and symptom irritability to more appropriately direct treatment decision-making for shoulder disorders.[2] However, this system is still in the conceptual stage and requires systematic evaluation, refinement, and validation before it can be recommended for clinical use.

The Theory and Research Literature Specific to the Topic

Geoffrey Maitland began promoting the concept of symptom irritability in the 1965[71] to determine the intensity for which examination and intervention procedures were prescribed.[9,16,72,73] Symptom irritability is defined as tissue readiness to accept physical stress.[8,14,15] It is important for a provider to be able to reliably determine the level of tissue readiness for physical stress as improper levels of physical stress applied to tissues can be detrimental to the patient.[10] However, symptom irritability has only been defined well enough for its measurement properties to be clearly evaluated for low back pain.[16]

The Physical Stress Theory (PST) postulates that tissues will adapt and remodel in a predictable manner based upon the stresses placed upon them.[10] When tissues are provided with physical stresses that are too low, the tissue will atrophy and this can lead to tissue death.[10] However, if physical stresses are too high, the tissue may experience rupture or tissue death.[10] Thus, in order to facilitate optimal patient outcomes, it is imperative for the provider to determine the amount of physical stress that will provide

www.manaraa.com

either maintenance of the tissue or hypertrophy of the tissue based upon the needs and goals of the plan of care.[10]

However, it is important that the determination of intervention intensity also must encompass factors such as movement and alignment factors, extrinsic factors, psychosocial factors, and physiological factors.[7,10] Movement and alignment factors include muscle performance, motor control, posture and alignment, pre-morbid physical activity level, and occupational and leisure activities.[7,10] Extrinsic factors include footwear, ergonomic environment, and gravity.[7,10] Psychosocial factors include kinesiophobia, catastrophization, depression, and anxiety.[74] Physiological factors include medications, age, systemic pathology, and obesity.[7,10]

As there are no clinical markers for the degree of inflammation available, and measurements such as erythrocyte sedimentation rate, C-reactive protein, and plasma viscosity testing are impractical in clinical situations, the construct of symptom irritability has been utilized by clinicians to determine the intensity of examination and intervention.[11,12] In order to encompass all of these aspects necessary to determine appropriate intensity, numerous researchers and clinicians have proposed criteria for which to base these diagnostic decisions (Table 2).[7-9,13,14]

www.manaraa.com

Table 2: Proposed Factors of Symptom Irritability

Ability to actively participate in the intervention[7]
Age[7,10]
Available support systems[7,10]
Body morphology[7,10]
Concurrent medications[7,10]
Home and workplace demands and requirements[7,10]
Level of disability[8,10,14]
Limb dominance[7]
Number of comorbidities[7,10]
Pain persistence (time it takes for the symptoms to subside once aggravated) [9,13]
Presence of wounds[7]
Psychosocial issues[7,8,10]
Severity of symptoms[9,13,14]
Stage of healing[7,8,10]
Static position tolerance[10,13]
Vigor of activity required to provoke a patient's symptoms[9,13,14]

A number of the factors listed in Table 2 have been vetted by a panel of experts with extensive clinical, research, and publication experience on shoulder pain to develop criteria to determine shoulder symptom irritability.[2] The shoulder symptom irritability classification system was initially proposed by Kelley et al in 2009[14] and was recommended for the classification of patients to aid in clinical decision-making for intervention intensity. The scale includes high, moderate, and low classification levels.[14] Classification was based on the following components: pain, mobility, and extent of disability.[14] It was suggested that those components would be conceptually defined by pain level, pain at night or at rest, disability scores, presence of pain at end range of motion (ROM), and the relationship of active ROM (AROM) compared to passive ROM (PROM).[14,27]

www.manaraa.com

Pilot data on the reliability of this shoulder symptom classification scheme was collected via a prospective, single-session, repeated measures design including patients with shoulder pain in an outpatient setting.[75] Eighteen subjects were assessed for shoulder pain irritability by two physical therapists who were blinded to each other's ratings. Inter-rater reliability was high (K=0.88-0.92, agreement 91.6-92.3%).[75] However, the study was significantly limited by the small sample size, inconsistency regarding pain assessment during ROM measurements, and arbitrary operationalization of the low, moderate and high cut-off scores for the patient reported outcome measures.[75] These limitations would need to be corrected in future studies in order to improve the internal validity and generalizability of the study outcomes.

Summary of What Is Known and Unknown About the Topic

Symptom irritability has been utilized extensively by clinicians and researchers for many years[9,15,16,61,63,72,73] and is proposed to aid in the determination of examination and intervention intensity, specifically for the shoulder.[2] There have been multiple methods of measurement of symptom irritability for various body regions[7,8,15,16] and one that has been vetted by a group of clinical and research experts that specifically relates to shoulder disorders.[2] And, while this shoulder symptom irritability classification system has not been appropriately defined to clearly determine reliability and validity, pilot data from a small study revealed an excellent trend toward good reliability.[75]

However, many aspects related to shoulder symptom irritability are yet unknown. The reliability and validity of the shoulder symptom irritability classification system have yet to be determined.[2] Secondly, the specific procedures to be matched with shoulder symptom irritability levels at operationally defined intensity levels are also unknown at

www.manaraa.com

this time.[2] Additionally, the usefulness of the shoulder symptom irritability classification system in the determination of appropriate intervention intensity is unknown, both from a clinical utility perspective and also from a cost/benefit perspective.[2]

The Contribution This Study Makes to the Field

This study provides a better understanding of the reliability of the shoulder symptom irritability classification system. Furthermore, it begins to build the necessary framework of correlation between diagnostic classification and treatment decision-making. Finally, it provides evidence of the importance of functional status in the symptom irritability classification system.

It is anticipated that the shoulder symptom irritability classification scale will be integrated with health condition (pathoanatomy) and body functions and structure (impairments) as per the STAR-Shoulder diagnostic classification system to appropriately prescribe rehabilitation intervention and reduce unwarranted variation in clinical practice.[2] Ultimately, the reduction in unwarranted variation is expected to result in reduced costs for the health care system and improved functional outcomes for patients.

Summary

It has been suggested that the shoulder is one of the most complex regions in the human body to diagnose because there is simultaneous movement of multiple joints, direct observation of movement can be obscured by muscle or adipose tissue, patient history is frequently vague, and there are a multitude of tests (clinical and imaging tests) that are not adequate to determine an accurate diagnosis.[76] Since diagnosis is one of the six major elements of patient management[77] and is a prerequisite for treatment[78] as it is

www.manaraa.com

necessary to select the appropriate intervention, facilitate communication between providers, and improve outcomes;[47,79] reliable labeling, or classification, is necessary to hone the intensity of interventions to most efficiently and effectively address the patient's problem.[80]

Pathologic classification has been insufficient to effectively direct treatment selection and intensity and thus it is important to include shoulder symptom irritability classification and physical impairment classification.[2] As physical impairment classification is not expected to be mutually exclusive, symptom irritability classification is utilized to provide clarity to the diagnostic process to aid in the determination of mutually exclusive diagnostic categories.

Shoulder symptom irritability is based on the principles of the PST.[2,10] There are many clinical indicators of shoulder symptom irritability of which, a panel of expert clinicians and researchers of shoulder pain, have been reduced to 5 separate factors.[2] Pilot data on the shoulder symptom classification system demonstrates a trend toward excellent reliability.[75] This study provides a better understanding of reliability and utility of the shoulder symptom irritability system.

www.manaraa.com

CHAPTER 3: METHODOLOGY

Introduction

This study employed a quasi-experimental observational design utilizing repeated measures (specific aim 1), followed by cross-sectional analysis (specific aims 2 and 3). The target sample size for the study was 25 providers and 90 patients. Patient-reported outcome measures were selected based on their reliability, validity, and internationally accepted use.[34,81,82] Given that there is no single universal patient-reported outcome measure for the shoulder, multiple measures were utilized during the third aim of this project.

Research Methods Employed

Experimental Design

For the reliability phase, the experimental design utilized repeated measures for inter-rater reliability. For the final two phases, a cross-sectional design was employed utilizing data gathered during the reliability phase. (Figure 1)

Specific Aim 1

The first specific aim was to determine the reliability of the shoulder symptom irritability classification system. To address this aim, we analyzed paired rater judgments of shoulder symptom irritability (high, moderate or low) from consecutive patients with shoulder pain. Raters were physical therapists from multiple sites trained in rating shoulder symptom irritability. Prevalence-adjusted, bias-adjusted Kappa for ordinal scales (PABAK-OS)[83] and percent agreement were the primary measures of reliability.

www.manaraa.com

Specific Aim 2

The second specific aim was to compare the level of functional limitation between shoulder symptom irritability groups. To address this aim, we analyzed patient-reported functional measures using analysis of variance with post-hoc analysis, in order to compare functional disability across different levels of shoulder symptom irritability. To preserve the validity of this analysis, only those subjects receiving the same classification by both raters were included. This methodology decreases the risk of confounders, as inter-rater reliability could otherwise affect the comparison of functional limitation between shoulder symptom irritability groups. The independent variable was the shoulder symptom irritability classification, and dependent variables included patient-reported functional status measures. The hypothesis was that patients with higher irritability would report greater functional deficits.

Specific Aim 3

The final specific aim was to determine if the level of shoulder symptom irritability dictates the chosen intervention intensity. To address this aim, raters selected intervention choices for each of the included patients, utilizing a pre-specified list of possible physical therapy interventions (Appendix H).

Data analysis included PABAK-OS for correlation and independent t-test for group differences. The hypothesis was that patients with high irritability would be prescribed interventions aimed at minimizing the physical stress to the affected tissue(s), while patients with low irritability would be prescribed interventions at a higher intensity to address the physical impairments.

www.manaraa.com

Specific Procedures Employed

Subjects

Requests for raters to participate in the study were sent to all 87 outpatient physical therapists in the St. Luke's University Health Network. The expected response rate was 25% due to the need for 2 raters at each site and participation interest in the study. Patient subjects were recruited from a convenience sample of consecutive patients presenting for physical therapy consultation for shoulder pain. As our pilot data demonstrated K>0.85 with similar methodology,[75] this study was powered at 80% to determine a K>0.80 with a sample size of 48 with a null K value of 0.40.[84] However, as only those subjects classified the same by both raters would be included in phase 2, there was the expectation of a significant drop in sample size between the first phase and the phase 2 of this study. Thus, doubling the required sample size was prudent to maintain the power of the subsequent analyses. Based on historical records, a patient sample size of at least 90 subjects was anticipated over a 6-month period.

Criteria for inclusion/exclusion

Rater Group

Inclusion criteria were state licensure as a physical therapist and regular clinical practice with patients with shoulder disorders, defined as a minimum of 500 clinical hours per year, and greater than or equal to 10% of caseload consisting of patients with shoulder disorders.

Exclusion criteria included not meeting inclusion criteria.

www.manaraa.com

Patient Group

Inclusion criteria were presenting with a chief complaint of shoulder pain, not extending to the neck, for outpatient physical therapy consultation.

Exclusion criteria included illiteracy in English and age less than 18 years. Additionally, subjects were excluded from the study if they presented with pain or symptoms distal to elbow, had shoulder surgery on the symptomatic side in the past year, if active or passive cervical spine ROM reproduced shoulder pain, had a positive Spurling's test, or if they were unable to complete the patient reported functional questionnaires. Subjects found to have a need for referral to another medical professional would have been provided with the appropriate referral. If the reason for referral was such that it would prevent them from participating safely in the study, that subject would have been excluded from subsequent testing.

Institutional Review Board (IRB) approval

Ethics approval has been obtained from the Institutional Review Boards of St. Luke's University Health Network (2016-61) and Nova Southeastern University (2016-379). Written informed consent was obtained from each subject prior to enrollment in the study.

Methods and Instrumentation

Instrumentation / Tests and Measures

Demographic information questionnaire

The survey (Appendix B) collected demographic data from raters including name, age, years of practice, type of advanced certification(s) held, gender, entry-level degree, and highest earned degree.

www.manaraa.com

Shoulder Symptom Irritability Classification System

Raters were asked to classify patient subjects in one of three shoulder symptom irritability classifications (Appendix G) based upon pain level, the presence of night or resting pain, the onset of pain during motion, differences between active and passive range of motion, and level of disability.[2,14,27]

Patient-Reported Outcome Scales

Three patient-rated outcome scales were administered for the purpose of enhancing generalizability, as there is no single gold standard patient-reported outcome scale accepted throughout the world for patients with shoulder pain.[34,35]

Focus On Therapeutic Outcomes (FOTO)

The FOTO functional scale (FS)[85] is a computerized adaptive test (CAT) and was administered via iPad (iPad 2, Apple, Cupertino, CA) at each clinic. The FOTO FS has been found to be a reliable and valid measurement system for outpatient orthopedic rehabilitation.[85-87] The FOTO FS was developed utilizing Item Response Theory and thus is a ratio scale that ranges from 0-100 with 0 being completely limited in all functional activity and 100 equated to full functional ability.[88-91]

In more recent studies, the FOTO questionnaire has demonstrated good construct validity and responsiveness for patients with shoulder complaints.[82,92] As a CAT, the FOTO questionnaire has a low burden on patients, with a mean test administration time of 1 minute and 29 seconds (SD = 90 seconds).[82] Furthermore, the standard error of the mean (SEM) has been found to be 1.30 with a minimal detectable change with 95% confidence (MDC_{95}) of 3.60-10.88 functional score units.[82,92] Minimal clinically important improvement (MCII) has been found to be 8 points utilizing receiver operating

www.manaraa.com

characteristic curve analysis.[82] When patients are grouped by quartile, the MCII is suggested to be 23, 10, 5, and 2 functional score change scores for the lowest through the highest quartiles upon intake, respectively.[82]

Additionally, FOTO was utilized to collect demographic data for each patient including comorbidities, age, gender, height, weight, chronicity of symptoms, type of insurance used, level of fear avoidance, and number of surgeries (Appendix D).

Penn Shoulder Score (PSS)

The Penn Shoulder Score (PSS), originally published in 1999[93] and validated in 2006,[94] is a self-report questionnaire consisting of three sections: pain, satisfaction, and function. The function subscale consists of twenty (20) items, each on a 4-point Likert scale (Appendix E). Each item is scored as 0 (can't do at all), 1 (much difficulty), 2 (with some difficulty), or 3 (no difficulty). The item scores are then summed to determine the subscale score out of 60 (no difficulty for all items). Resultant scores for each subscale are divided by the total range from 0-100 with 0 as greatest disability and 100 as no disability.[93]

The PSS has demonstrated good test-retest reliability ($ICC_{2,1}$ = 0.94) with a SEM_{90} of 8.5.[94] The MDC_{90} is 12.1, and the minimal clinically important difference (MCID) was found to be 11.4.[94]

American Shoulder and Elbow Surgeons (ASES) Shoulder Score

The American Shoulder and Elbow Surgeons (ASES) Shoulder Score, originally published in 1994[95] and validated in 2002,[96] is a self-report questionnaire consisting of two sections: one visual analog scale (VAS) to measure pain, and ten items to measure activities of daily living. The questionnaire takes 3 minutes to complete and is scored as

www.manaraa.com

follows: [(10 – VAS pain) x 5] + (5/3 x sum of ADL items).[97] As the items in the PSS are identical to the ASES, the PSS form was enhanced with 1 additional question to obtain both PSS scores and ASES scores with minimal responder burden (Appendix E). Resultant scores for each subscale range from 0-100 with 0 as greatest disability and 100 as no disability.[97]

The ASES has demonstrated good to excellent test-retest reliability (ICC = 0.61-0.96) with an SEM of 6.7.[97] The MDC_{95} is 11.2,[96] and the MCID was found to be 12.0.[98] Furthermore, a recent systematic review found the ASES to be one of the only patient-reported functional scales for rotator cuff disease to have measurement error below 10% of the global score.[99]

Numeric Pain Rating Scale

The Numeric Pain Rating Scale (NPRS) is an 11-point Likert scale that can be used to measure pain intensity. The NPRS is a standard pain assessment scale that uses a 0-10 scale (no pain to worst pain imaginable, respectively) to determine a patient's level of pain. Patients rate their level of pain in the last 24 hours. The NPRS has demonstrated good reliability ($ICC_{2,1}$=0.74) and responsiveness (MDC = 2.5, MCID = 1.1) in subjects with shoulder pain[100] and excellent reliability in an upper extremity orthopaedic population.[101] Furthermore, the NPRS has been used to assess pain severity of both traumatic and atraumatic etiologies.[102]

Range of Motion

Measurements of active range of motion (AROM) are performed to determine limitations in motion, and the impact of movement on symptoms. Active flexion of the shoulder is performed in an upright position. Care was taken to ensure the patient

www.manaraa.com

maintains an upright position throughout the examination and during subsequent follow-up examinations. All passive movements of the shoulder were performed in the supine position.[103-106]

All methods are moderately correlated with more definitive radiographic and 3D kinematic measurements. Goniometric measurements of shoulder AROM in symptomatic patients demonstrates fair-good reliability with regards to intra- and inter-rater reliability (Inter-rater Rho = 0.64-0.80; Intra-rater Rho = 0.53-0.91).[103-106] Passive range of motion (PROM) demonstrates even greater reliability with intra-examiner ICC values = 0.98, and inter-examiner ICC values ranging from 0.87-0.89.[104]

In order to measure flexion AROM, the patient is positioned in a standing position and is asked to actively flex the shoulder to end range.[27] ROM is measured by placing the axis of the goniometer on the greater tuberosity. The stationary arm is aligned with the midline of the trunk. The movable arm is aligned with the lateral epicondyle.[27]

To measure flexion PROM, the patient is positioned in supine with the arm comfortably by the side. The examiner passively flexes the shoulder until end range is reached (with no compensatory movements from the thorax and the lumbar spine). ROM is measured by placing the axis of the goniometer on the greater tuberosity. The stationary arm is aligned with the midline of the trunk. The movable arm is aligned with the lateral epicondyle.[27]

End Feel - Pain

End feel is generally defined as the sensation perceived by the clinician when resistance to motion is felt, and whether that sensation of resistance is due to pain or tissue tension.[40] In this study, end feel was specifically used to indicate the onset of pain

www.manaraa.com

in relation to onset of tissue resistance. While the ability to utilize end feel to determine sequence of pain in relation to tissue resistance has demonstrated variable inter-rater reliability (K = 0.62 to 0.76,[41] K_w = -0.01 to 0.70[40]), it has generally shown good intra-rater reliability (K = 0.48 to 0.59,[41] K_w = 0.59 to 0.87[40]) and is frequently used for clinical decision-making.[40]

Procedures

See Appendix A for the flow chart of study procedures. Raters were recruited via email and personal request. A minimum of 2 raters at each site was required for enrollment in the study. Raters were consented in person and demographic data on the raters was collected (Appendix B). The raters were then trained with the following materials: (1) Collaborative Institutional Training Initiative (CITI) training for those involved in consenting patients, (2) the reading of the Staged Approach for Rehabilitation Classification: Shoulder Disorders[2] with direction to pay special attention to the section on Level 3 classification and Table 3,[2(pp 795-6)] and (3) a short online narrated presentation to reinforce understanding of the content (https://youtu.be/a-QiJ5-bKKQ).[107] The intent of this training method was to increase the generalizability of the study results and to avoid overly specialized training methods that would be difficult to reproduce clinically.

Consecutive patients were recruited by the raters from their regular caseload (Appendix C). Patients received a brief explanation of the study, provided informed consent, and were asked to complete the functional questionnaires (Appendices D-E) as part of the outpatient admissions process. Included in the admissions process, FOTO was also utilized to collect demographic data from subjects including comorbidities, age, gender, height, weight, chronicity of symptoms, and level of fear avoidance. The first

www.manaraa.com

therapist rated the patient's shoulder symptom irritability classification during the normal examination process utilizing the intake forms (Appendices F-G). After the first rater completed their examination and prior to any intervention that may have changed the shoulder symptom irritability, a second rater, blinded from the first rater's assessment, then examined and rated the subject (Appendices F-G). In addition to the shoulder symptom irritability rating, both raters were asked to provide a treatment intensity recommendation based on the examination findings (Appendices H-I). Data collection forms were placed in a sealed security-tint envelope and sent via interoffice mail for data entry and analysis. Data were collected from December 1, 2016-June 9, 2017. All data were entered and maintained on a secure, password-protected server (RedCap, Nashville, TN; https://redcap.slhn.org/).

Statistical Analysis

IBM SPSS Statistics (Version 25) was utilized to perform all statistical analysis. Descriptive statistics were used to characterize both raters and patients. Frequencies were utilized for categorical variables and means with standard deviations for continuous variables.

A repeated measures design, utilizing two raters per subject, was utilized to determine inter-rater reliability. The raters independently rated the subject's shoulder symptom irritability level utilizing the shoulder symptom irritability classification system.[2,14] The inter-rater reliability was evaluated using the prevalence-adjusted, bias-adjusted Kappa for ordinal scales statistic (PABAK-OS).[83,84] For evaluation of statistical significance, a two-tailed confidence interval was utilized with α set to 0.05, and the null hypothesis was that the PABAK-OS is <0.40.[84] The PABAK-OS statistic was selected

www.manaraa.com

due to the inherent unequal distribution of irritability levels in clinical practice, and to minimize the effect of any rater bias.

The concept of shoulder symptom irritability is hypothesized to impact treatment decision making.[2,14] Treatments directed at a patient with high shoulder symptom irritability should include those to minimize physical stress and modify symptoms whereas the interventions directed toward a patient with low irritability should include moderate to high physical stress and be specifically directed at addressing the patient's physical impairments.[2] In the intermediate between high and low is the moderate shoulder symptom irritability group which would receive interventions with mild to moderate physical stress addressing basic functional activity restoration and beginning to address the patient's physical impairments.[2] If a clinician incorrectly classifies a patient as high instead of moderate shoulder symptom irritability, the patient would receive interventions to minimize physical stress and modify symptoms to facilitate addressing the physical impairments on a subsequent encounter. Thus, minimal time is lost, and the patient would still benefit from this incorrect dosing of treatment as symptoms in the moderate shoulder symptom irritability group are still limiting basic daily functional tasks. However, if a patient was misclassified as high instead of low shoulder symptom irritability, the treatments would not be addressing the physical impairments and would instead be focused on minimizing symptoms that do not significantly need to be modified to facilitate the improvement of those underlying physical impairments.

Consider a patient with subacromial pain with intermittent resting pain, moderate pain at rest, pain at end range of motion, and can do basic functional tasks despite mild discomfort with heavier tasks. If the clinician misclassifies the patient with high shoulder

www.manaraa.com

symptom irritability instead of moderate shoulder symptom irritability, the initial intervention would likely include activity modification to decrease pain with activity, thermal modalities and pharmacological analgesics as needed to reduce the pain at rest, and monitor physical impairments for accurate planning of future encounters. This patient would have missed the opportunity to begin working on restoring basic functional activities but would have learned to alter basic activities to decrease symptoms, which in turn, would have facilitated the restoration of those basic functional activities naturally as symptoms decreased. Additionally, the patient would have missed the opportunity to begin addressing the specific physical impairments associated that may be associated with subacromial pain such as passive mobility deficits, poor motor control, and muscular weakness. However, the treatment for a patient with high irritability would focus on encouraging the use of unaffected regions which would help to facilitate neuromuscular patterning and minimize atrophy and further loss of mobility and strength. Thus, only minimal time would have been lost from this misclassification.

However, if the patient with subacromial pain had no pain at rest or at night, low amounts of pain throughout the day, and only with higher level activities was classified as having high shoulder symptom irritability instead of low shoulder symptom irritability, the initial intervention would likely include activity modification to decrease pain with activity, thermal modalities and pharmacological analgesics as needed to reduce the pain at rest, and monitor physical impairments for accurate planning of future encounters. This patient would have missed the opportunity to begin addressing the patient's physical impairments and begin working on restoring the higher-demand functional activities which are limited. Treatments for a patient with high irritability, directed at reducing

www.manaraa.com

symptoms at rest and basic activity modification, would not only be needless, but are a waste of time and a threat to therapeutic alliance, and thus may significantly delay recovery.[108]

Thus, it was determined that near misses of one level would be weighted 2/3. The level of 2/3 was chosen specifically because a near miss means the clinician would likely result in a treatment that partially addresses the patient's problem at an intensity that will still improve their condition in a significant manner and cause only a slight delay in facilitating recovery, whereas a miss by two levels would likely result in treatment that is of low therapeutic value to the patient and would likely delay the patient's recovery.

PABAK-OS was calculated utilizing the web-based application from Single Case Research.[109] To complement a thorough description of reliability, percent agreement was also reported. For evaluation of statistical significance, a two-tailed confidence interval was utilized with α set to 0.05, and the null hypothesis was that the PABAK-OS was <0.40.[84]

Analysis of variance with post-hoc analysis was utilized for evaluation of differences in patient-reported functional limitation and pain subscales between shoulder symptom irritability groups. Furthermore, ordinal regression was utilized to determine the strength of the pain and functional subscales to predict SSIC. For evaluation of statistical significance, α was set to 0.05.

Lastly, to evaluate the correlation between intervention intensity and diagnosed classification of shoulder symptom irritability, the PABAK-OS statistic was used. Additionally, independent t-tests were utilized to evaluate for differences between groups

www.manaraa.com

for hypotheses 1 and 2 of aim 3. For evaluation of statistical significance, α was set to 0.05.

Formats for Presenting Results

Presentation

I plan to initially present this to the larger physical therapy community at the Combined Sections Meeting (CSM). I will be submitting for platform presentation at CSM 2019.

Publication

This dissertation would split well into three papers: one for reliability (Specific Aim 1), one for correlation with treatment selections (Specific Aim 2), and one for comparison of the degree of functional limitation between shoulder symptom irritability groups (Specific Aim 3).

Since this specific dissertation is focused on rehabilitation professionals, specifically physical therapists, my first target would be Physical Therapy as it has the greatest reach and impact factor for the target audience. If denied from that journal, my next targeted journal would be to obtain the largest orthopedic physical therapy audience with Journal of Orthopaedic and Sports Physical Therapy. Lastly, if denied from both of those journals, I would focus on Archives of Physical Medicine and Rehabilitation due to the relatively high impact factor and focus on rehabilitation professionals.

Resources Used

Grant Awards

A grant was obtained from the Auxiliary of the St. Luke's University Hospital for $5,000. Grant funding was utilized in the following manner.

www.manaraa.com

Support Personnel:

$960 for a data technician [2 hr./wk. x 24 weeks x $20/hr.].

Subject Recruitment

$1,010 for $10 gift card for each subject recruited [$10 x 101 subjects]

$2,020 for $10 gift card for each subject recruited to each rater [$10 x 2 x 101 subjects]

$1,010 for $10 gift card for each subject recruited to each clinic coordinator [$10 x 101 subjects]

Additional Funding

The following was paid for via self-funding and/or employer funding.

Transportation/Registration/Room/Board for Presentation of Results

$2,000 for printing, transportation, registration, room, and board for presentation of results at the 2019 Combined Sections Meeting of the American Physical Therapy Association.

Data Analysis:

$200 IBM SPSS Statistics Premium Version 24, 24-month license

Equipment & Supplies:

$100 for general supplies (e.g., paper, copies, pens, internet access)

IRB Submission:

$3,500 for initial review with St. Luke's University Health Network

$950 for final report with St. Luke's University Health Network

Total Cost:

The total funding resources utilized to complete this study was $11,750.

www.manaraa.com

CHAPTER 4: RESULTS

Introduction

Thirty-six (36) physical therapists from 16 sites completed rater training. Of those trained, 24 raters from 11 sites submitted data meeting the inclusion criteria for the study. One-hundred-one (101) patients were included in the data analysis, and 6 were excluded for the following reasons: pain or symptoms distal to the elbow (1), failure to complete forms (1), cervical spine involvement (3), and history of ipsilateral upper extremity surgery (1) (Figure 2).

Figure 2: Participant inclusion in the analysis for this study

Findings

Participant Demographics

Initially, we received responses from and trained 35 raters from 16 sites. Of the 24 raters that submitted patient data for this study, the mean age was 33.9 (+/-7.3) years with a mean of 8.1 (+/-6.7) years of experience in clinical practice. Females accounted for 41.7% of the raters, 87.5% had earned a DPT or higher, and 54.2% had earned American Board of Physical Therapy Specialties (ABPTS) Certification in either Sports or Orthopaedic Physical Therapy. As the samples did not fall within a normal curve due

www.manaraa.com

to positive skewness for age and years of practice (Figures 3-4), the scale data for those two variables were compared utilizing a Mann-Whitney U test.[110] Nominal data were compared utilizing Chi-square testing as long as the cell counts were >5; and in the case of highest earned degree, Fisher's Exact test was utilized as 50% of the cell counts were fewer than 5. There were no statistically significant demographic differences between those raters that submitted patient data and those that did not submit data. (Table 3)

Table 3: Rater demographics

	Participated (n=24)	Did not participate (n=13)	p
Age (mean, SD)*	33.9 (7.3)	33.5 (7.5)	0.76
Years of Practice (mean, SD)*	8.1 (6.7)	7.4 (6.2)	0.60
Female (n, %)‡	10 (41.7%)	5 (38.5%)	0.85
Entry-Level Degree of DPT (n, %)‡	17 (70.8%)	8 (61.5%)	0.56
Highest Earned Degree of DPT (n, %)†	21 (87.5%)	10 (76.9%)	0.64
ABPTS Certification‡	13 (54.2%)	3 (23.1%)	0.07

Data was compared via: * Mann-Whitney U, ‡ Chi-Square test, † Fisher's Exact test

www.manaraa.com

Figure 3: Rater Age Distribution

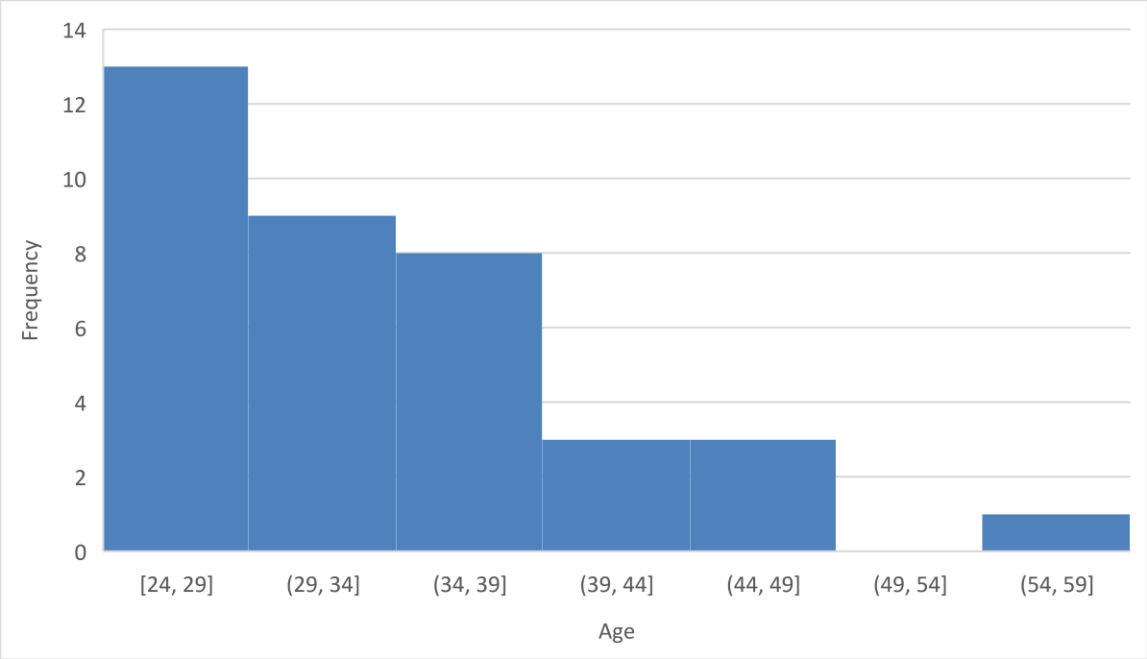
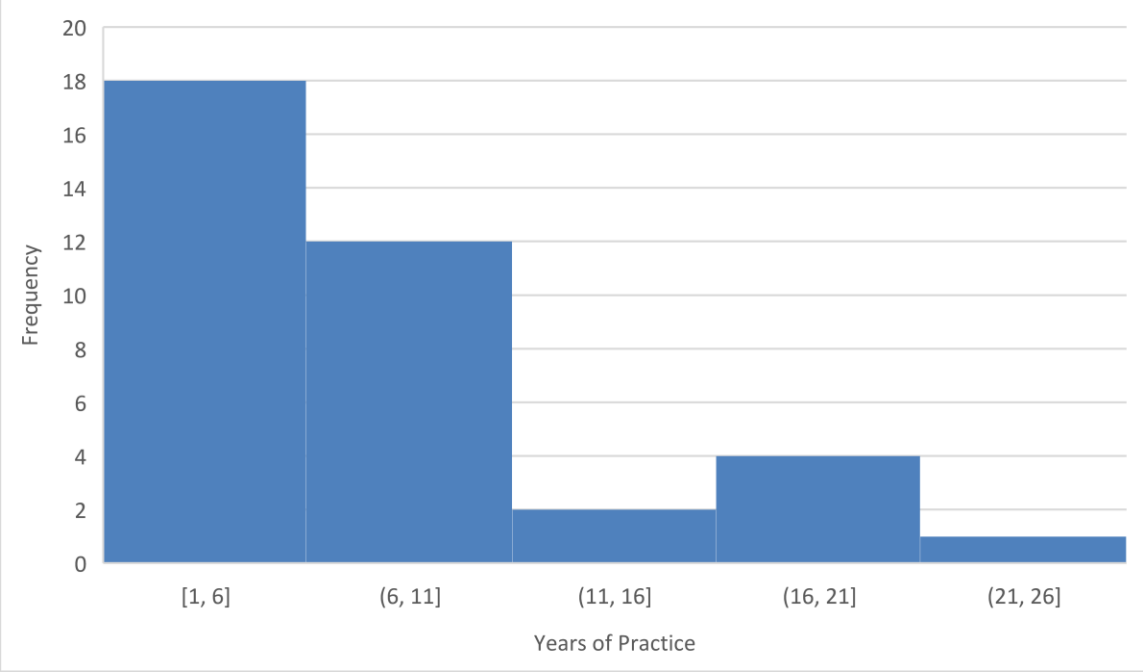


Figure 4: Rater Years of Practice Distribution



www.manaraa.com

The mean age of the 101 consecutive patients included in the study analyses was 56.0 (+/-16.0) years, with females accounting for 65.3% of the sample. The majority of the sample (88.1%) was right hand dominant and 56.4% of the entire sample had complaints of right shoulder pain. Additionally, 43.6% of the sample had elevated levels of fear avoidance. Only 21.8% of the patients presented with acute pain of fewer than 3 weeks duration. The largest single group of patients presented with subacute pain of 3 weeks to 3 months (37.6%). Finally, 14.9% of patients presented with pain that had lasted 3-6 months and 25.7% of patients presented with pain of greater than 6 months duration (Table 4).

Table 4: Patient demographics

	n = 101
Age (mean, SD)	56.0 (16.0)
Female (n, %)	66 (65.3%)
Right hand dominance (n, %)	89 (88.1%)
Right arm affected (n, %)	57 (56.4%)
Elevated fear avoidance (n, %)	44 (43.6%)
Acuity (time since onset of symptoms)	
0-7 days	5 (5.0%)
8-14 days	7 (6.9%)
15-21 days	10 (9.9%)
22-90 days	38 (37.6%)
91 days to 6 months	15 (14.9%)
Over 6 months	26 (25.7%)

Aim 1: Reliability of Shoulder Symptom Irritability Classification

All 101 subjects were included in the inter-rater reliability analysis. As anticipated, the SSIC with the greatest frequency of selection was moderate (46.3%) followed by a relatively even distribution of 28.4% and 25.4% for low and high irritability, respectively (Table 5). The inter-rater reliability of the shoulder symptom

www.manaraa.com

irritability classification system was PABAK-OS = 0.69 (95% Confidence Interval = 0.59-0.78) and the percent agreement between raters was 68% (Table 6).

Table 5: Frequency of Shoulder Symptom Irritability Rating

SSIC Ratings	n (%)
Low Irritability	38 (28.4%)
Moderate Irritability	62 (46.3%)
High Irritability	34 (25.4%)

SSIC = Shoulder Symptom Irritability Classification

Table 6: Inter-Rater Reliability of Shoulder Symptom Irritability Classification – All Sites

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	23	9	0
	Moderate	5	30	6
	High	1	12	15

PABAK-OS = 0.69 (95% CI 0.60, 0.78)
Rater Agreement = 67%

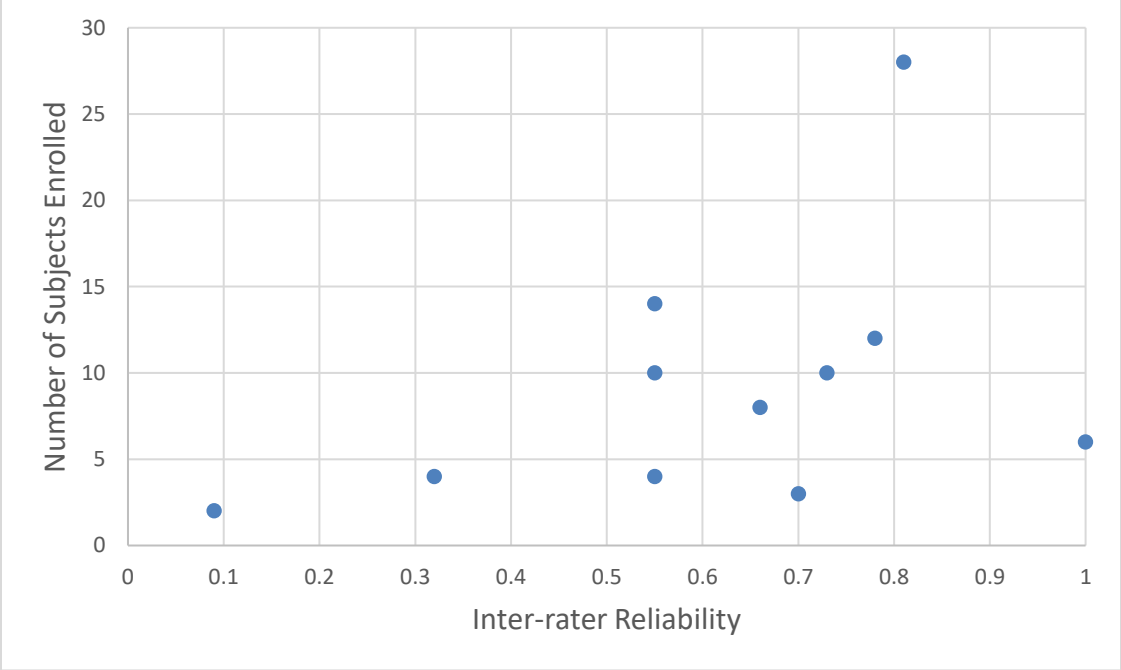
Within each of the 11 participating sites, inter-rater reliability ranged from PABAK-OS = 0.09-1.0 and rater agreement ranged from 0-100% and is summarized in Table 7 (contingency tables for sites can be found in Appendix J). One potential limitation was that raters may learn each other's rating habits if blinding was not maintained. Thus, inter-rater reliability was also assessed separately for those sites that submitted 10 or more subjects and those that submitted less than 10 subjects. To assess the risk of unblinding, sites were grouped into those that submitted 10 or more subjects to be analyzed in the study (increased inherent risk of discussion and inter-rater learning) and those that submitted fewer than 10 subjects to be analyzed (increased inherent risk of discussion and inter-rater learning). The threshold of 10 or more subjects was chosen as the two groups would have a nearly equal number of sites (5 vs. 6) as seen in Figure 5.

www.manaraa.com

Table 7: Inter-Rater Reliability of Shoulder Symptom Irritability Classification by Site

Site	n	PABAK-OS (95% CI)	Rater Agreement
1	6	0.55 (0.25, 0.85)	70%
2	12	0.55 (0.29, 0.80)	50%
3	10	0.78 (0.50, 1.0)	75%
4	6	0.66 (0.33, 1.0)	63%
5	14	0.09 (0, 0.77)	0%
6	2	0.81 (0.63, 0.99)	79%
7	7	0.73 (0.43, 1.0)	70%
8	7	1.0 (0.61, 1.0)	100%
9	3	0.70 (0.15, 1.0)	67%
10	4	0.55 (0.08, 1.0)	50%
11	26	0.32 (0, 0.80)	25%

Figure 5: Inter-Rater Reliability Compared by Site



Sites that submitted 10 or more subjects had an inter-rater reliability of PABAK-OS = 0.71 (95% CI 0.60-0.82), and sites that submitted fewer than 10 subjects had an inter-rater reliability of PABAK-OS = 0.63 (95% CI 0.45, 0.82) (Tables 8-9). Thus, there was no significant improvement or degradation of reliability between groups that

www.manaraa.com

have had increased experience rating subjects when compared to those groups that have had less experience, but the same degree of training.

Table 8: Inter-Rater Reliability of Shoulder Symptom Irritability Classification - Sites with n < 10

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	6	2	0
	Moderate	1	6	4
	High	0	4	4

PABAK-OS = 0.63 (95% CI 0.45, 0.82)
Rater Agreement = 59%

Table 9: Inter-Rater Reliability of Shoulder Symptom Irritability Classification - Sites with n ≥ 10

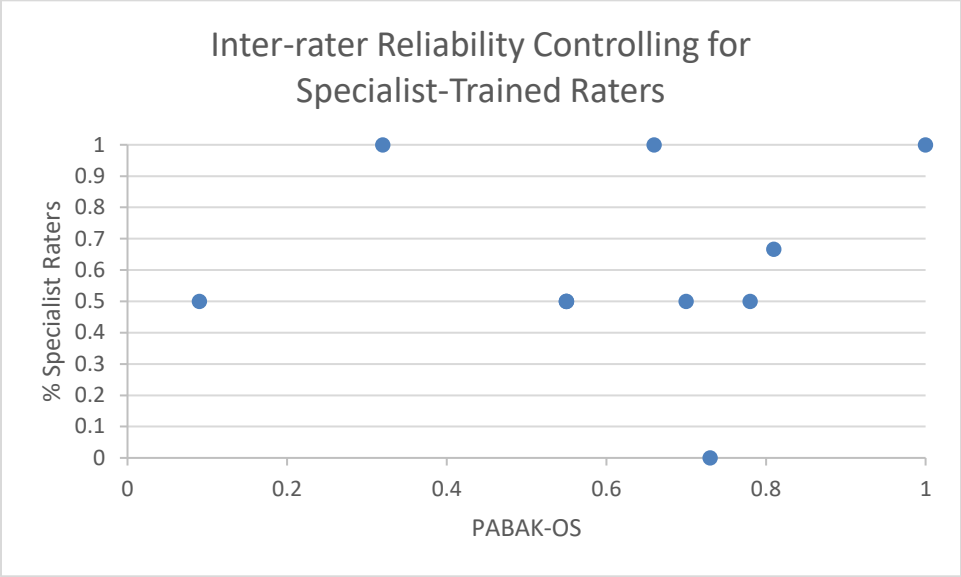
		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	17	7	0
	Moderate	4	24	2
	High	1	8	11

PABAK-OS = 0.71 (95% CI 0.60, 0.82)
Rater Agreement = 70%

Additionally, another potential limitation was that the number of specialists in the site may positively influence the reliability of the ratings. However, when the ratio of specialists was compared between the two sites with the greatest reliability to the two sites with the worst reliability, there was no trend discovered. The two sites with the greatest inter-rater reliability consisted of 80% specialist raters, whereas the two sites with the lowest inter-rater reliability consisted of 75% specialist raters (Figure 6).

www.manaraa.com

Figure 6: Inter-rater reliability controlling for specialist-trained raters

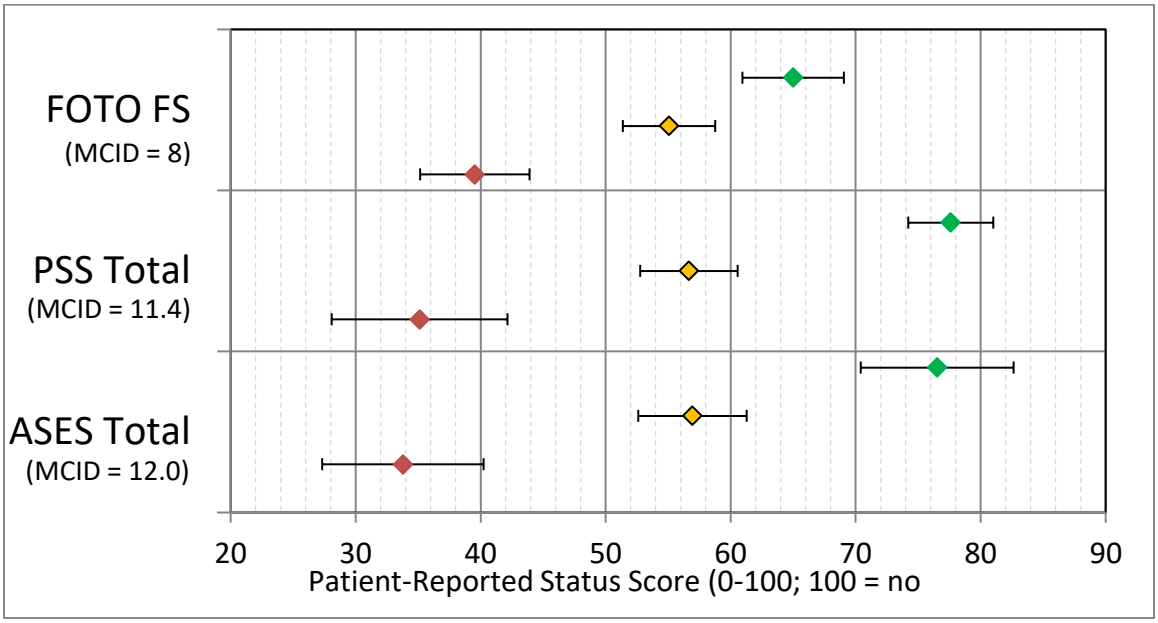


Aim 2: Compare level of functional limitation between irritability groups

Raters agreed upon the rating of 68 patients, and subsequently, those 68 subjects were included in the analysis of the second aim. One subject did not complete the FOTO FS and thus the analyses of FOTO FS only included a sample size of 67. Mean functional limitation scores with 95% confidence intervals for each of the shoulder symptom irritability groups are depicted for all three PRO measures in Figure 7. Analysis of variance demonstrated significant differences in functional limitation between irritability groups for the PSS, ASES, and FOTO FS ($p<0.001$) as described in Table 10. Furthermore, Bonferroni post hoc analysis demonstrated significant differences between all groups for all patient-reported functional measures, with an effect size of functional score on shoulder symptom irritability group ranging from 3.20-6.80 (Table 11).

www.manaraa.com

Figure 7: Functional limitation differences between shoulder symptom irritability groups



PSS, Penn Shoulder Score. ASES, American Shoulder & Elbow Surgeons Score. FOTO FS, Focus on Therapeutic Outcomes Functional Score. Error bars = 95% Confidence Interval

Red = High shoulder symptom irritability

Yellow = Moderate shoulder symptom irritability

Green = Low shoulder symptom irritability

Table 10: Functional limitation differences between shoulder symptom irritability groups

Patient-reported functional measure	n	df	F	p
PSS Total	68	2	67.38	<0.001
PSS Function	68	2	45.62	<0.001
PSS Pain	68	2	63.68	<0.001
ASES Total	68	2	45.27	<0.001
ASES Function	68	2	34.11	<0.001
ASES Pain	68	2	26.20	<0.001
FOTO FS	67	2	29.06	<0.001

PSS, Penn Shoulder Score. ASES, American Shoulder & Elbow Surgeons Score. FOTO FS, Focus on Therapeutic Outcomes Functional Score.

www.manaraa.com

Table 11: Effect size of functional limitation between shoulder symptom irritability groups

Patient-reported functional measure	Low to Moderate	Moderate to High
PSS Total	6.80 (4.33-9.25)	6.12 (3.67-8.58)
PSS Function	5.81 (3.34-8.26)	4.82 (2.35-7.27)
PSS Pain	5.89 (3.43-8.34)	6.66 (4.21-9.13)
ASES Total	5.19 (2.74-7.66)	5.39 (2.93-7.85)
ASES Function	5.45 (2.99-7.91)	3.67 (1.21-6.13)
ASES Pain	3.20 (0.75-5.67)	4.78 (2.32-7.24)
FOTO FS	3.53 (1.08-5.99)	4.86 (2.39-7.31)

Effect size (95% confidence interval)

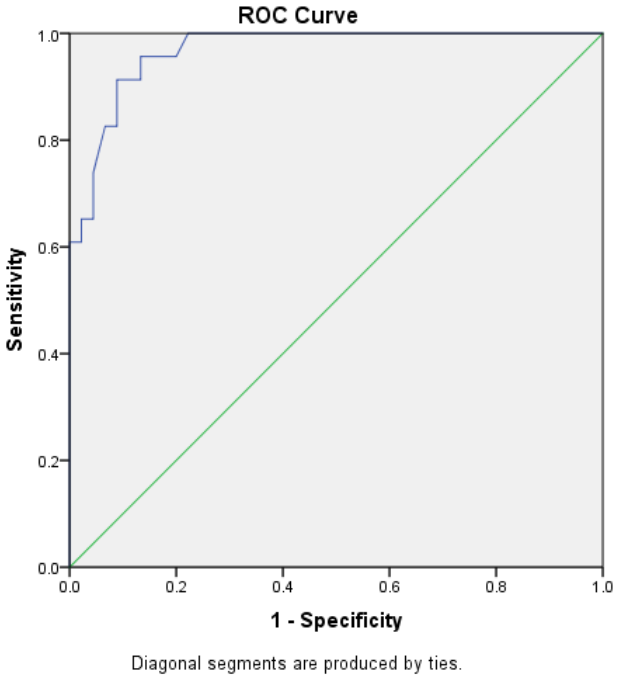
Additional exploratory analyses were completed as an extension of the study's aims due to the large differences in functional scores between shoulder symptom irritability classification groups. To the author's knowledge, there have been no studies that have determined cut-off scores based upon severity of functional limitation to aid in the selection of shoulder symptom irritability.[2] Thus, receiver operating characteristic curve analyses for the different patient-reported functional outcome scales and subscales were used to determine the cut-off values that would maximize the sensitivity and specificity of each scale (Figures 8-21).

The optimal cut-off scores to discriminate high from moderate shoulder symptom irritability and low from moderate shoulder symptom irritability, along with their respective sensitivity and specificity values, are shown in Table 12. The ROC curves were produced utilizing the 68 pairs of rater data from the 68 subjects with matched rater classifications. Cut-off scores are summarized in Table 12. These cut-offs were compared with rater classification of shoulder symptom irritability to determine the percent agreement as a measure of the reliability of the cut-off scores to complement the sensitivity and specificity derived from the ROC curve analysis (Figures 22-24). The cut-off scores with the best agreement, as shown in Table 13, were the PSS Function

www.manaraa.com

Subscale (79%), PSS Total Score (78%), ASES Total Score (78%), and PSS Pain Subscale (72%).

Figure 8: Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for PSS Total Score



www.manaraa.com

Figure 9: Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for PSS Total Score

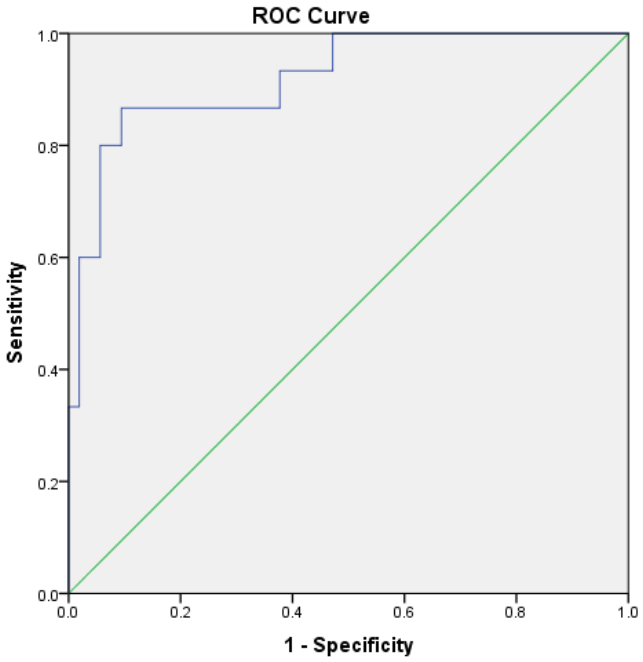
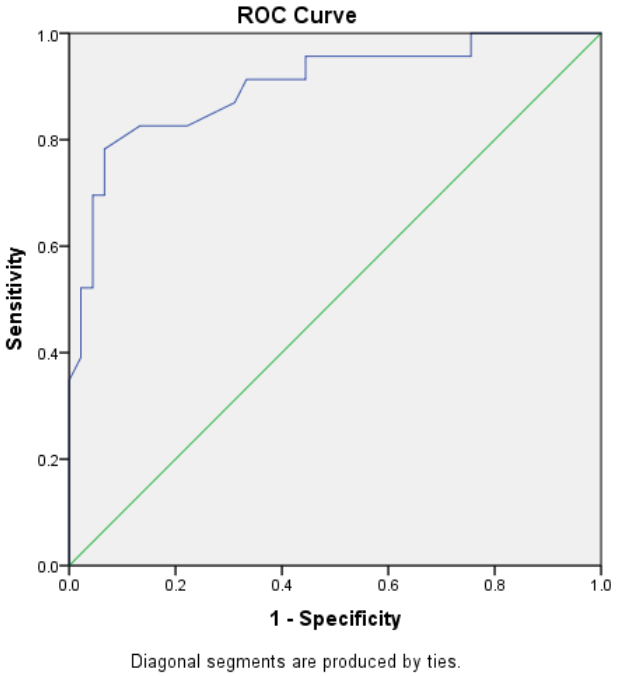


Figure 10: Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for ASES Total Score



www.manaraa.com

Figure 11: Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for ASES Total Score

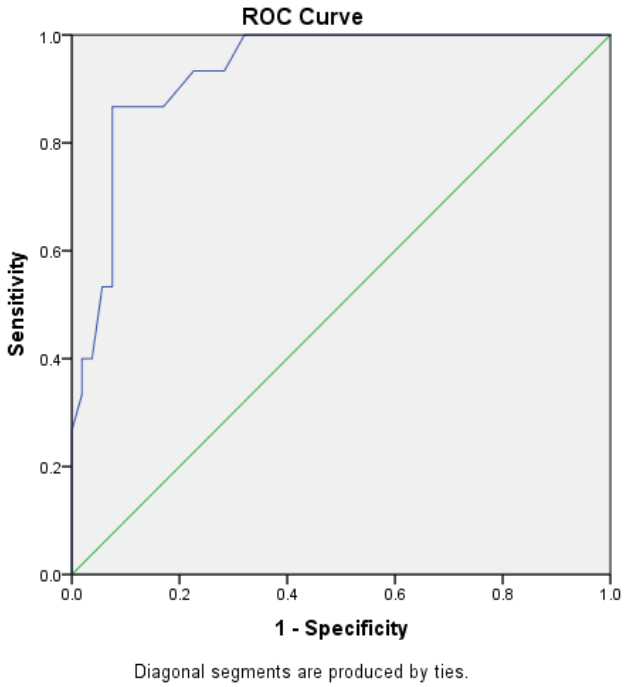
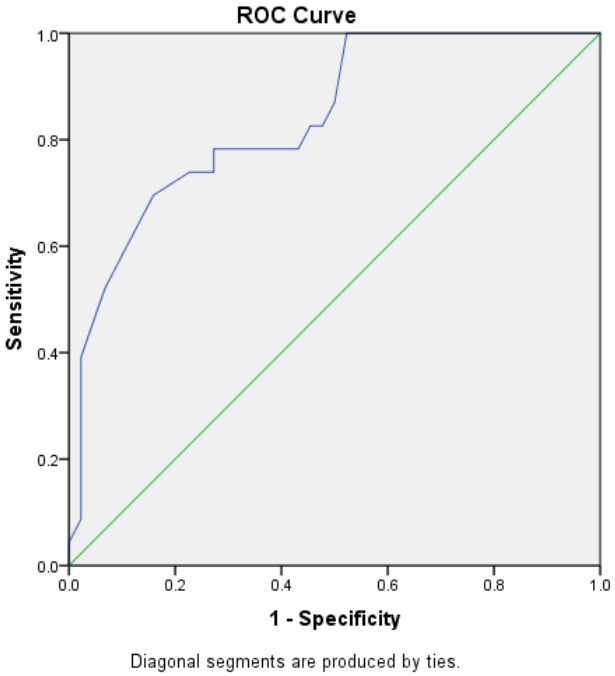


Figure 12: Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for FOTO FS



www.manaraa.com

Figure 13: Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for FOTO FS

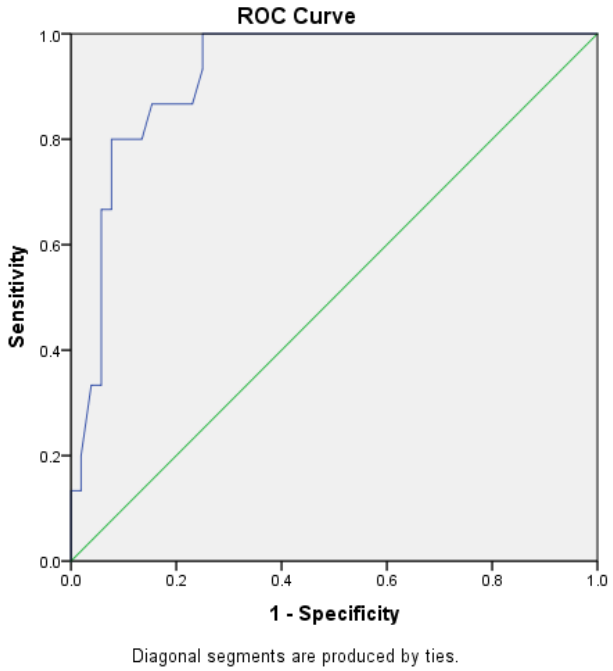
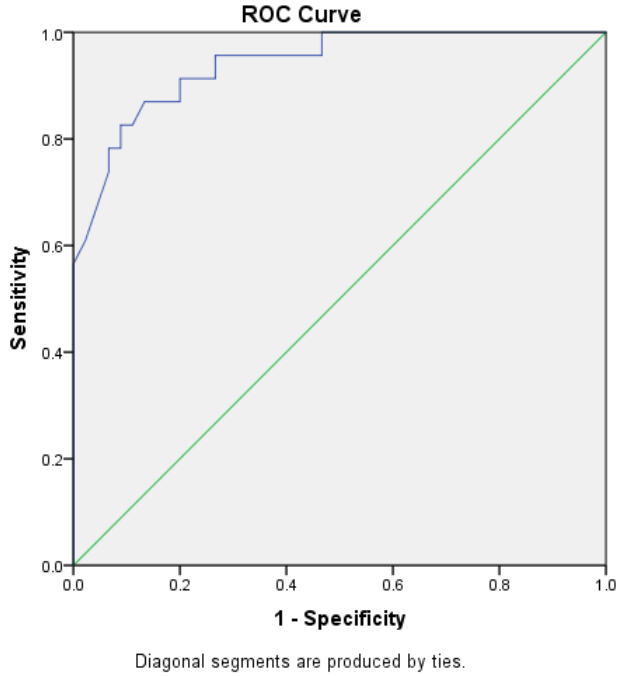


Figure 14: Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for PSS Function Subscale



www.manaraa.com

Figure 15: Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for PSS Function Subscale

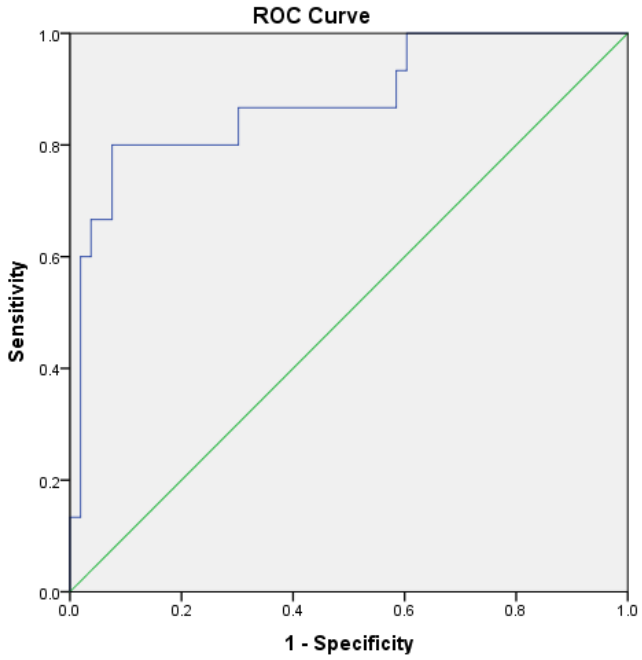
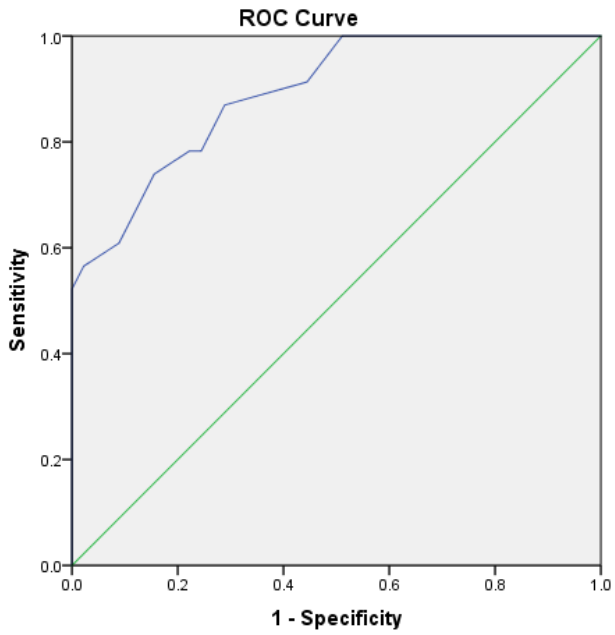


Figure 16: Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for ASES Function Subscale



www.manaraa.com

Figure 17: Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for ASES Function Subscale

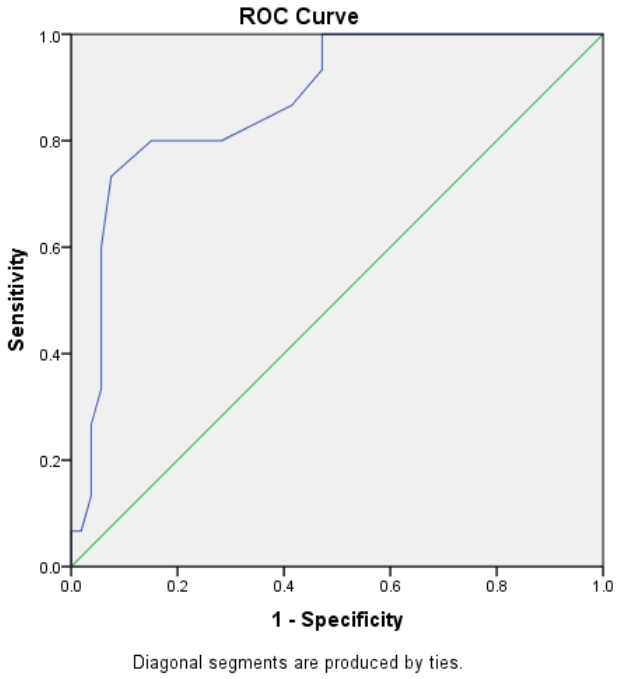
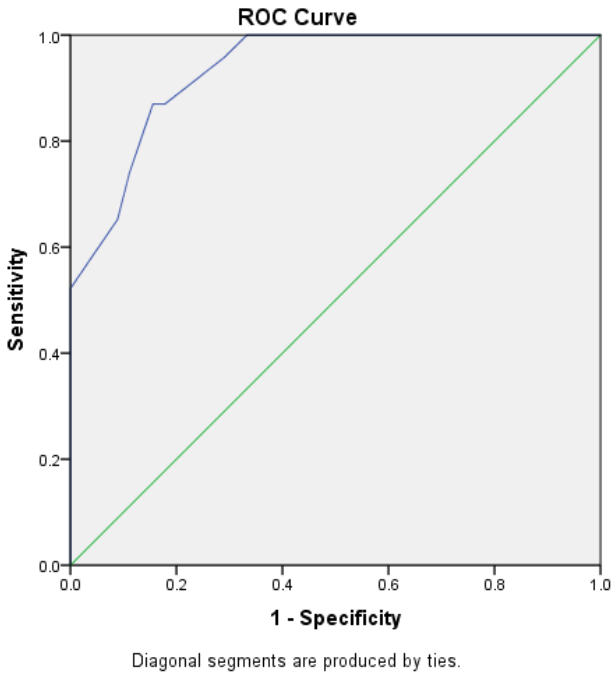


Figure 18: Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for PSS Pain Subscale



www.manaraa.com

Figure 19: Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for PSS Pain Subscale

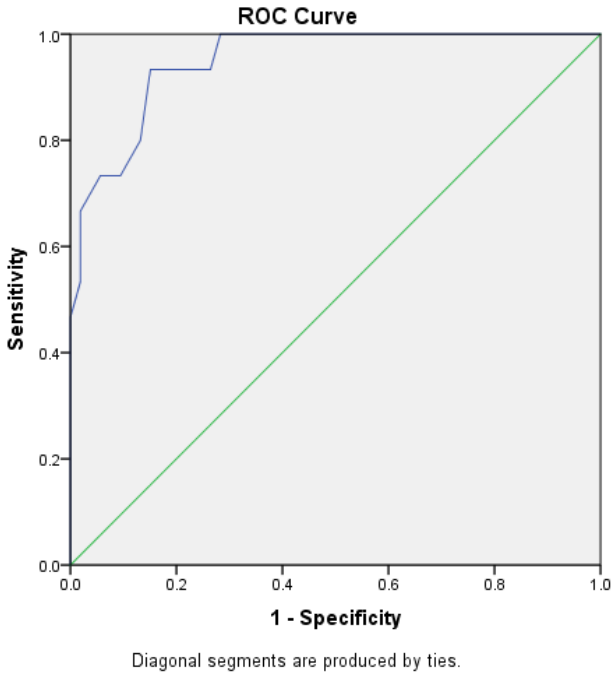
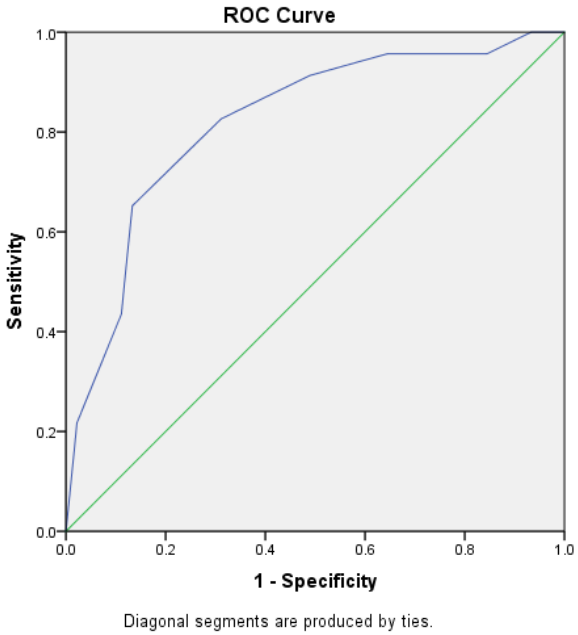


Figure 20: Receiver operating characteristic curve for low vs. moderate/high shoulder symptom irritability for ASES Pain Subscale



www.manaraa.com

Figure 21: Receiver operating characteristic curve for high vs. moderate/low shoulder symptom irritability for ASES Pain Subscale

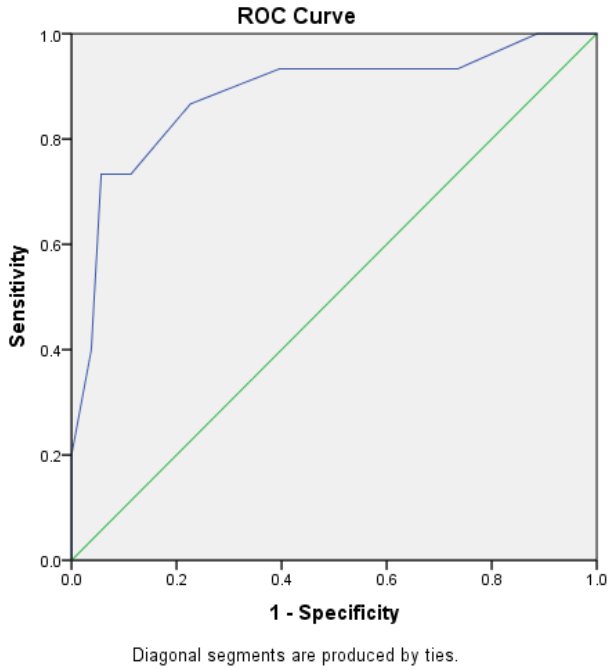


Table 12: Receiver operating characteristic curve results

Scale	High Irr (≤)	Sn	Sp	Low Irr (≥)	Sn	Sp
PSS Total	47.9	.867	.906	68.6	.913	.911
ASES Total	48.3	.867	.925	65.8	.826	.867
FOTO	47.0	.800	.923	62.0	.696	.841
PSS Function	27.9	.800	.925	43.7	.870	.867
ASES Function	22.5	.800	.849	32.5	.739	.844
PSS Pain	15.5	.933	.849	20.5	.870	.844
ASES Pain	27.5	.867	.774	32.5	.826	.689

High Irr, Cut-off to differentiate high shoulder symptom irritability from moderate and low shoulder symptom irritability; Low Irr, Cut-off to differentiate low shoulder symptom irritability from moderate and high shoulder symptom irritability; PSS Pain, Penn Shoulder Scale Pain Subscale; PSS Function, Penn Shoulder Scale Function Subscale; PSS Total, Penn Shoulder Scale Total Score; ASES Pain, American Shoulder and Elbow Surgeons Pain Subscale; ASES Function, American Shoulder and Elbow Surgeons Function Subscale; ASES Total, American Shoulder and Elbow Surgeons Total Score; FOTO, Focus on Therapeutic Outcomes Functional Score

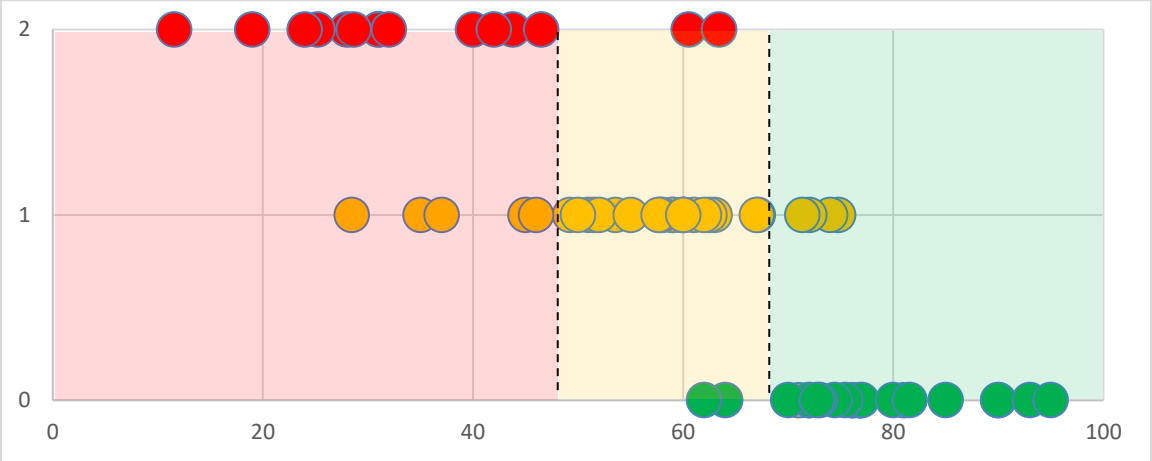
www.manaraa.com

Table 13: Patient-reported outcome measure cut-off scores

Irritability	High	Moderate	Low	Both Raters Agree	All Ratings
PSS Total	0 - 47.9	48.0 - 68.5	68.6 - 100	.78	.68
ASES Total	0 - 48.3	48.4 - 65.7	65.8 - 100	.78	.66
FOTO	0 - 47.0	47.1 - 61.9	62.0 - 100	.66	.61
PSS Function	0 - 27.9	28.0 - 43.6	43.7 - 60	.79	.69
ASES Function	0 - 22.5	22.6 - 32.4	32.5 - 50	.65	.59
PSS Pain	0 - 15.5	15.6 - 20.4	20.5 - 30	.72	.62
ASES Pain	0 - 27.5	27.6 - 32.4	32.5 - 50	.57	.50

PSS Pain, Penn Shoulder Scale Pain Subscale; PSS Function, Penn Shoulder Scale Function Subscale; PSS Total, Penn Shoulder Scale Total Score; ASES Pain, American Shoulder and Elbow Surgeons Pain Subscale; ASES Function, American Shoulder and Elbow Surgeons Function Subscale; ASES Total, American Shoulder and Elbow Surgeons Total Score; FOTO, Focus on Therapeutic Outcomes Functional Score

Figure 22: Distribution of PSS total score based upon rater selection



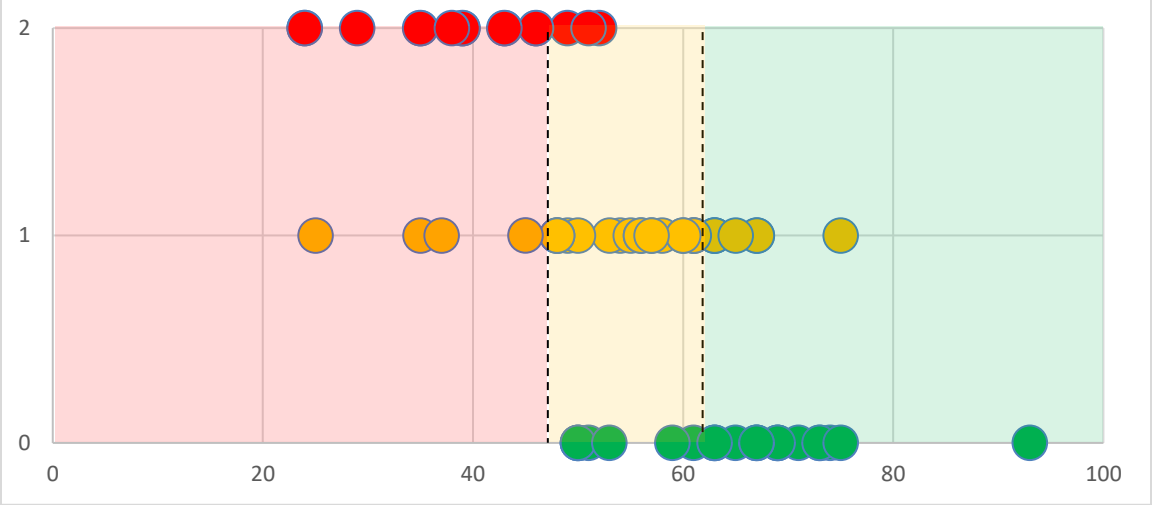
PSS, Penn Shoulder Score
Red Markers = Rater selected high shoulder symptom irritability
Yellow Markers = Rater selected moderate shoulder symptom irritability
Green Markers = Rater selected low shoulder symptom irritability
Red Shaded area = Cut-off score determination of high shoulder symptom irritability
Yellow Shaded area = Cut-off score determination of moderate shoulder symptom irritability
Green Shaded area = Cut-off score determination of low shoulder symptom irritability

www.manaraa.com

Figure 23: Distribution of ASES total score based upon rater selection

ASES, American Shoulder and Elbow Surgeons Score
Red Markers = Rater selected high shoulder symptom irritability
Yellow Markers = Rater selected moderate shoulder symptom irritability
Green Markers = Rater selected low shoulder symptom irritability
Red Shaded area = Cut-off score determination of high shoulder symptom irritability
Yellow Shaded area = Cut-off score determination of moderate shoulder symptom irritability
Green Shaded area = Cut-off score determination of low shoulder symptom irritability

Figure 24: Distribution of FOTO Functional Score based upon rater selection



FOTO, Focus on Therapeutic Outcomes
Red Markers = Rater selected high shoulder symptom irritability
Yellow Markers = Rater selected moderate shoulder symptom irritability
Green Markers = Rater selected low shoulder symptom irritability
Red Shaded area = Cut-off score determination of high shoulder symptom irritability
Yellow Shaded area = Cut-off score determination of moderate shoulder symptom irritability
Green Shaded area = Cut-off score determination of low shoulder symptom irritability

www.manaraa.com

To decrease the risk of sample bias, these derived cut-off scores were then compared to all rater choices (n=202, except for FOTO FS n= 200). The agreement of the cut-scores ranged from 0.50 (ASES Pain) to 0.69 (PSS Function) as summarized in Table 13. In this secondary analysis, the cut-off scores with the best agreement were the PSS Function Subscale (69%), PSS Total Score (68%), ASES Total Score (66%), and PSS Pain Subscale (62%).

To determine the strength of influence different pain subscales and specific question items have on the determination of shoulder symptom irritability, ordinal regression was utilized to compare the two pain subscales from the PSS and ASES. It was hypothesized that the PSS Pain Subscale, which is a composite of 3 items, would have a greater relationship with shoulder symptom irritability groups than the ASES Pain Subscale. As summarized in Table 14, the PSS Pain subscale significantly influenced the selection of Shoulder Symptom Irritability Classification, while the ASES Pain Subscale did not.

Table 14: Parameter Estimates for Ordinal Regression of PSS Pain Subscale and ASES Pain Subscale

	Estimate	Standard Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
PSS Pain	-.531	.116	20.791	1	<.001	-.759	-.303
ASES Pain	-.017	.037	.203	1	.652	-.090	.056

PSS Pain, Penn Shoulder Score Pain Subscale; ASES Pain, American Shoulder and Elbow Surgeons Score Pain Subscale

Additionally, as clinicians are very busy and frequently do not feel they have time to ask multiple pain questions, it would be useful to understand the influence of specific items within the PSS Pain Subscale to more efficiently make treatment decisions. Thus,

www.manaraa.com

each of the 3 items within the PSS Pain Subscale was analyzed utilizing ordinal regression. As summarized in Table 15, the two items of the PSS Pain Subscale items that involve how pain influences function, "pain with normal activities (eating, dressing, bathing)" and "pain with strenuous activities (reaching, lifting, pushing, pulling, throwing)", were influential upon the classification of shoulder symptom irritability while the remaining item, "pain at rest with your arm by your side" was not found to be influential.

Table 15: Parameter Estimates for Ordinal Regression of PSS Pain Subscale items

	Estimate	Standard Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Pain at Side	.392	.238	2.709	1	.100	-.075	.859
Pain ADL	.844	.244	11.951	1	.001	.366	1.323
Pain Strenuous	.382	.194	3.863	1	.049	.001	.763

Pain at Side, pain at rest with your arm by your side; Pain ADL, pain with normal activities (eating, dressing, bathing); Pain Strenuous, pain with strenuous activities (reaching, lifting, pushing, pulling, throwing)

To determine the strength of influence pain subscales and functional subscales have in the determination of shoulder symptom irritability, ordinal regression was performed with shoulder symptom irritability as the dependent variable, and Penn Shoulder Score (PSS) Pain Subscale and Function Subscale as independent variables. Parameter estimates were -0.44 (95% CI -0.66, -0.22) for the pain subscale and -0.12 (95% CI -0.21, -0.04) for the function subscale (Table 16) demonstrating superior influence of the pain subscale on shoulder symptom irritability when compared to the function subscale. Additionally, the correlation between PSS Pain Subscale and Shoulder

www.manaraa.com

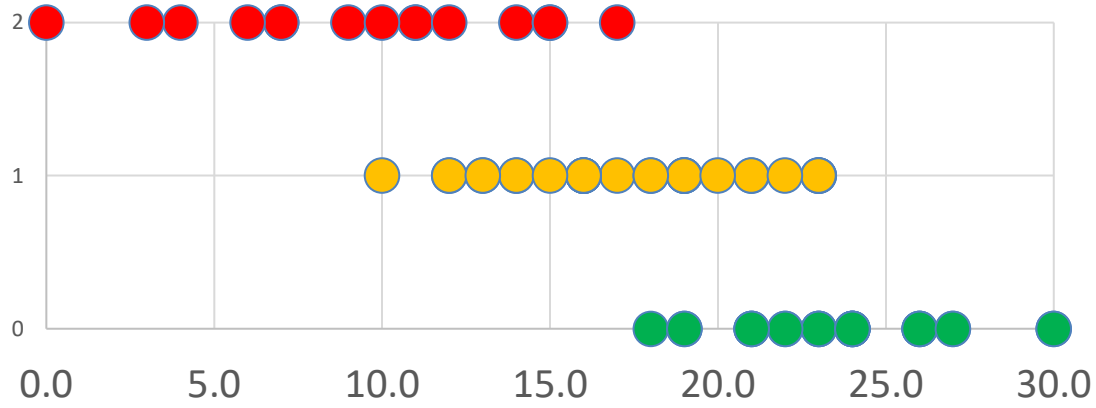
Symptom Irritability Classification groups was stronger than between PSS Function Subscale and Shoulder Symptom Irritability Classification groups (Figures 25-26).

Table 16: Parameter Estimates for Ordinal Regression of Pain and Function

	Estimate	Standard Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
PSS Pain	-0.439	0.114	14.732	1	<0.001	-0.663	-0.215
PSS Function	-0.123	0.045	7.584	1	0.006	-0.210	-0.035

PSS Pain, Penn Shoulder Score Pain Subscale; PSS Function, Penn Shoulder Score Function Subscale

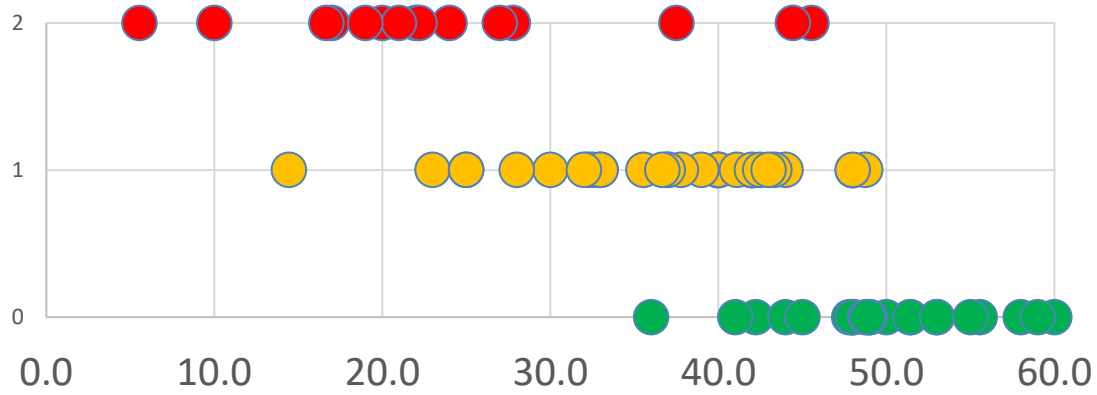
Figure 25: Correlation between PSS Pain Subscale and Shoulder Symptom Irritability



PSS, Penn Shoulder Score
Red Markers = Rater selected high shoulder symptom irritability
Yellow Markers = Rater selected moderate shoulder symptom irritability
Green Markers = Rater selected low shoulder symptom irritability

www.manaraa.com

Figure 26: Correlation between PSS Function Subscale and Shoulder Symptom Irritability



PSS, Penn Shoulder Score
Red Markers = Rater selected high shoulder symptom irritability
Yellow Markers = Rater selected moderate shoulder symptom irritability
Green Markers = Rater selected low shoulder symptom irritability

Aim 3: Determine if the level of irritability logically guides the chosen intervention

Raters were asked to select the shoulder symptom irritability classification that best described the patient and to select the treatment strategy best suited for the patient on that date of service. As shoulder symptom irritability is designed to determine treatment strategy and intensity, rater agreement between shoulder symptom irritability and treatment strategy was determined utilizing PABAK-OS and percent agreement. The sample size was 202 for this analysis (decision-making for each rater for each of the 101 patients). PABAK-OS was 0.82 (95% CI 0.75, 0.88) with 80% agreement (Table 17). Thus, the hypothesis that the level of shoulder symptom irritability is moderately correlated (K>0.40 and agreement>50%) with planned intervention intensity is accepted.

www.manaraa.com

Table 17: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	35	25	1
	Moderate	0	86	6
	High	0	7	42

PABAK-OS = 0.82 (95% CI 0.75, 0.88)
Rater Agreement = 81%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Additionally, there were no significant differences in agreement between shoulder symptom irritability and treatment strategy when dichotomized between specialist and non-specialist groups (p=0.56). However, there was a trend toward better agreement in the non-specialist group with PABAK-OS of 0.85 (95% CI 0.76, 0.95) when compared to the agreement of the specialist group with PABAK-OS of 0.79 (95% CI 0.69, 0.88) (Tables 18-19).

Table 18: Agreement of Specialists between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	17	16	1
	Moderate	0	44	3
	High	0	3	21

PABAK-OS = 0.79 (95% CI 0.69, 0.88)
Rater Agreement = 78%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Table 19: Agreement of Non-Specialists between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	18	9	0
	Moderate	0	42	3
	High	0	4	21

PABAK-OS = 0.85 (95% CI 0.76, 0.95)
Rater Agreement = 84%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

As there was a trend, but no significant differences regarding the originally selected measure of expertise as previously noted, additional metrics were investigated to determine if there were any factors likely to improve or diminish the matching of treatment strategy to SSIC. Other metrics of expertise that have been utilized in the literature are years of experience.[111-113] The American Physical Therapy Association denotes those with 5 or fewer years of experience as "New Professionals."[114] Thus, raters were dichotomized into those with more than 5 years of experience and those with less than or equal to 5 years of experience. As summarized in Tables 20-21, a trend was noticed that those with more than 5 years of experience had a greater likelihood of matching selected shoulder symptom irritability to treatment strategy, but no significant differences were found with 95% confidence.

www.manaraa.com

Table 20: Agreement of Raters with more than 5 years of experience between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	20	10	0
	Moderate	0	46	4
	High	0	4	22

PABAK-OS = 0.85 (95% CI 0.76, 0.94)
Rater Agreement = 83%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Table 21: Agreement of Raters with 5 or fewer years of experience between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	15	15	1
	Moderate	0	40	2
	High	0	3	20

PABAK-OS = 0.78 (95% CI 0.69, 0.88)
Rater Agreement = 78%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

If years of experience were a likely factor to improve the ability of the provider to select a matched treatment strategy, it would be logical that those clinicians with greater than 10 years of experience would have an even better agreement than those in the group with only greater than 5 years of experience. However, while the sample size is too small to be conclusive, this is not the case as summarized in Table 22, and when compared to those clinicians with 10 or fewer years of experience (Table 23), the trend was no longer present.

www.manaraa.com

Table 22: Agreement of Raters with more than 10 years of experience between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	13	9	0
	Moderate	0	36	3
	High	0	4	15

PABAK-OS = 0.82 (95% CI 0.71, 0.93)
Rater Agreement = 80%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Table 23: Agreement of Raters with 10 or fewer years of experience between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	22	16	1
	Moderate	0	50	3
	High	0	3	27

PABAK-OS = 0.82 (95% CI 0.73, 0.90)
Rater Agreement = 81%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

While the confidence interval of matched treatments described in Table 17 was small, the presence of any outliers that either improved or diminished the overall group's agreement in matching treatment strategy to SSIC was a concern and thus all raters' reliability are summarized in Table 24 and individual contingency tables for each rater can be found in Appendix K. PABAK-OS ranged from 0.10 to 1.0 across all raters. One rater had PABAK-OS of 0.1 which is more than 21 times the standard error from the mean overall agreement. However, even with removing this single rater, the overall agreement did not change even a single percentage point indicating that this single outlier did not have a significant impact on the overall agreement.

www.manaraa.com

Table 24: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

Rater	n	PABAK-OS (95% CI)	Matching
1	6	0.70 (0.31, 1.00)	67%
2	12	1.00 (0.73, 1.00)	100%
3	10	0.73 (0.43, 1.00)	80%
4	6	0.55 (0.16, 0.94)	50%
5	14	1.00 (0.75, 1.00)	100%
6	2	0.55 (0, 1.00)	50%
7	7	0.74 (0.39, 1.00)	71%
8	7	0.87 (0.51, 1.00)	86%
9	3	1.00 (0.45, 1.00)	100%
10	4	1.00 (0.53, 1.00)	100%
11	26	0.86 (0.68, 1.00)	85%
12	14	0.68 (0.53, 0.93)	79%
13	2	0.10 (0, 0.77)	0%
14	24	0.78 (0.58, 0.97)	75%
15	4	1.00 (0.53, 1.00)	100%
16	4	0.55 (0.08, 1.00)	50%
17	12	0.62 (0.35, 0.90)	58%
18	3	1.00 (0.45, 1.00)	100%
19	6	0.85 (0.46, 1.00)	83%
20	6	1.00 (0.61, 1.00)	100%
21	10	0.91 (0.61, 1.00)	90%
22	4	0.55 (0.08, 1.0)	50%
23	8	1.00 (0.67, 1.00)	100%
24	8	0.90 (0.58, 1.00)	89%

Furthermore, four other raters had PABAK-OS values of 0.55, which is more than 7 times the standard error from the mean overall agreement. However, after removing all five of these outliers (rater 4, 6, 13, 16, and 22), there was still not a significant shift in PABAK-OS (0.84; 95% CI 0.77, 0.92) to explain any differences in reliability due to a single outlier rater (Table 25). Also, to avoid a single clinic with good reliability and one of the largest contributors to the dataset from influencing the overall agreement of matching treatment strategy to SSIC, the raters in site 6 were removed from the dataset and it was re-analyzed for agreement. As summarized in Table 26, there was only a

www.manaraa.com

small, statistically insignificant shift in agreement to PABAK-OS = 0.81 (95% CI 0.73, 0.89). Thus, can be concluded that it is unlikely that any outliers in this study contributed significantly to improving or diminishing the overall group's agreement in matching treatment strategy to SSIC.

Table 25: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy – Outliers removed

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	33	19	1
	Moderate	0	82	5
	High	0	4	40

PABAK-OS = 0.84 (95% CI 0.77, 0.92)
Rater Agreement = 84%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Table 26: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy – Largest site removed

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	21	17	1
	Moderate	0	62	6
	High	0	5	34

PABAK-OS = 0.81 (95% CI 0.73, 0.89)
Rater Agreement = 80%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Summary of Results

Twenty-four clinicians rated a total of 101 patients who were included in the study. Inter-rater reliability of the Shoulder Symptom Irritability Classification system was 0.69 with no improvements in rating noted in sites submitting 10 or more patients. Significant differences were found between shoulder symptom irritability groups regarding functional limitation ($p<0.001$). The PSS Pain Subscale had a stronger influence over the classification of shoulder symptom irritability than the PSS Function Subscale. Raters selected "matched" treatment strategies 80% of the time, with no significant difference between raters who are specialists and those who are non-specialists.

www.manaraa.com

CHAPTER 5: DISCUSSION

Introduction

In this chapter, we will be discussing the results of the study and the implications of those results. First, the results suggest that the Shoulder Symptom Irritability Classification (SSIC) system is a reliable classification method. However, while the SSIC system does demonstrate good reliability, numerous questions arose regarding rater expertise, rater blinding and the findings of poorer reliability than was found during pilot testing.

Secondly, this study also determined that there is a significant difference in functional limitation between SSIC groups and is the first study, to our knowledge, to determine cut-off scores in patient-reported functional limitation outcome questionnaires to aid in the determination of shoulder symptom irritability. The results further demonstrate that even though function is strongly correlated with SSIC, pain is a stronger determinant of SSIC.

Lastly, the SSIC appears to greatly influence the prescription of overall treatment strategy. Furthermore, the correlation between matched treatment and SSIC did not significantly differ between expert and non-expert provider groups, indicating that the SSIC system has the same degree of validity across both groups of providers.

Discussion

Participant Demographics

The participants for this study represented a sample of convenience of 24 raters from 11 different sites. As we initially trained 35 raters from 16 sites, a comparative analysis was performed to determine if demographics of the raters contributed to non-

www.manaraa.com

submission of patient data. There was no significant difference between the demographics of those raters that were trained but did not participate and those that were trained and participated in the study. Thus, there is low risk that those who dropped out would likely have demonstrated poorer, or greater, agreement in rating or treatment strategy.

However, the sample of clinicians in the regional health network where the study was performed had a much higher frequency of having a DPT (87.5%) and being ABPTS certified (54.2%) when compared to the APTA's member demographic average of 44.4% DPT[115] and 7.9% ABPTS certification.[115,116] Thus, while further work is necessary to determine generalizability outside of this sample, it is promising that there was no difference in the frequency of specialists among those sites with better or worse inter-rater reliability, nor was there a difference in treatment strategy agreement in trained raters based upon specialty certification.

The patient demographics of our sample are representative of the patient population with shoulder pain. The patients in the present study had a mean age of 56.0 ± 16.0 and 65.3% of them were women, which is comparable to prevalence studies that have found 57% of patients with shoulder pain are women with a median age range of 55-64 years.[117] Additionally, arm laterality is similar to population-based studies with approximately 90% prevalence of right-hand dominant people[118] which is comparable to 88.1% found in the present study. Lastly, the majority of our sample had symptoms for >3 weeks which is similar to former studies.[119]

www.manaraa.com

Aim 1: Reliability of Shoulder Symptom Irritability Classification

The inter-rater reliability of the shoulder symptom irritability classification system is good[120] with a PABAK-OS of 0.69 (95% CI 0.59, 0.78). The null hypothesis set forth in this study that PABAK-OS <0.40 cannot be supported with 95% confidence. However, the alternative hypothesis that PABAK-OS is >0.60 also cannot be supported with 95% confidence. These results are slightly lower than those found in a pilot study with similar methodology to this study.[75] One possible explanation for these lower results was that the pilot study[75] improperly assigned arbitrary cutoff scores to the patient-reported functional limitation outcome measures. The pilot study by Kareha et al,[75] utilized arbitrary cutoff scores to aid in the clinical decision-making process, but as there was no clinical rationale or research base to the decision, it may have elevated the inter-rater reliability scores.

Other well accepted and commonly utilized scales have demonstrated similar or worse inter-rater reliability. The treatment-based classification algorithm for the low back pain has been found to have inter-rater reliability of K = 0.52 (95% CI = 0.27, 0.77)[24] in one study and only slightly better in an earlier study of K = 0.60 (95% CI 0.56, 0.64).[121] Neck pain classification has been found to have a very high inter-rater reliability (K = 0.95, 95% CI 0.87, 1.0), but the rating of this system was based only upon documented information and was not performed in real-time.[66]

The McKenzie classification system has been analyzed for inter-rater reliability on multiple occasions. Different studies have contested the validity of previous studies assessing inter-rater reliability based upon training level and varying criterion measures. Utilizing only the three main classifications and raters highly trained in the McKenzie

www.manaraa.com

system, subsequent studies have found inter-rater reliability with K = 0.70 (95% CI 0.45, 0.96),[122] K = 0.64 (95% CI 0.18, 1.0),[123] and K = 0.84 (95% CI 0.62, 1.0).[124] However, the two largest reliability studies found poor inter-rater reliability (K = 0.26; 95% CI 0.20, 0.32)[125] and (K = 0.37-0.44)[126] based upon varying levels of formal McKenzie training.

Maitland's musculoskeletal pain irritability system demonstrated a prevalence-adjusted, bias-adjusted Kappa (PABAK) of 0.50 (95% CI 0.26, 0.74).[16] Lastly, the classification of scapular dyskinesis has long been a component of physical examination[42] but even the most reliable classification system for scapular dyskinesis was found to have K_w between 0.48 and 0.61.[127]

While the null hypothesis number of 0.40 was chosen specifically to determine if the result was not worthy of use in clinical practice,[84] the test hypothesis threshold of 0.60 was chosen simply due to the arbitrary threshold proposed by Landis and Koch.[120] However, Sim and Wright suggested that arbitrary thresholds should not be utilized as the Kappa statistic is a continuum that is limited by constraints of the population.[84] Thus, while the inter-rater reliability found in this study does not support the testing hypothesis of this study with 95% confidence, it is similar, if not better than, many widely accepted classification systems and can be considered sufficient reliability for clinical use.[24]

A potential consideration for reliability studies is the possibility that individual sites or raters may dramatically skew the final results. When investigating the reliability of individual sites, three sites stood out as significant outliers as their reliability was greater than 3 times the standard error, indicated by Portney and Watkins[128] as the

www.manaraa.com

reference for extreme values, from the mean PABAK-OS. The individual site inter-rater reliability of the three sites were PABAK-OS scores of 0.09, 0.32, and 1.0, respectively.

As mentioned earlier, inter-rater reliability for the entire dataset was PABAK-OS of 0.69 (95% CI 0.59, 0.78). However, analysis after removing the three outlier sites did not significantly change inter-rater reliability as PABAK-OS was 0.69 (95% CI 0.59, 0.79). Furthermore, with only removing the high-end outlier PABAK-OS only moved insignificantly to 0.67 (95% CI 0.57, 0.77), and with only removing the low-end outliers PABAK-OS also only moved insignificantly to 0.71 (95% CI 0.61, 0.81).

One site did submit a very large number of subjects to the study which also may have skewed the results. Site 6 (Table 7) submitted 28 subjects (27.7% of all subjects) over the course of the study with a PABAK-OS of 0.81 (95% CI 0.63, 0.99). PABAK-OS analysis excluding this data resulted in a reliability of 0.64 (95% CI 0.53, 0.75). Thus, it can be concluded that while there was a shift in reliability when the largest site was removed, it was not statistically significant as the confidence intervals overlap. Additionally, none of the individual sites significantly influenced the overall results of the study, nor when removing the sites with the 5 largest subject contributors, increasing the generalizability of these results.

A consideration and possible limitation of this study design was the influence of repeated performance of the classification resulting in the possibility of rater unblinding. Raters were specifically told to avoid discussing ratings throughout the duration of the study to maintain blinding. However, to assess the risk of unblinding, sites were grouped into those that submitted 10 or more subjects to be analyzed in the study (increased inherent risk of discussion and inter-rater learning) and those that submitted fewer than

www.manaraa.com

10 subjects to be analyzed (increased inherent risk of discussion and inter-rater learning). No significant differences were found between the groups as 95% confidence intervals of both groups overlapped. Thus, there is no significant improvement or degradation of reliability between groups that have had increased experience rating subjects when compared to those groups that have had less experience, but the same degree of training, indicating there is a low risk that unblinding occurred.

Although repetitive of use of this skill is able to possibly reduce the performance of a measure, perhaps the opposite can also be true. As with any cognitive skill, disuse has been shown to decrease the performance of that skill.[129] However, no conclusions can be convincingly drawn regarding the impact of the delayed use of the rating system after training based upon this data as confidence intervals overlap. However, given the trend toward worse reliability with longer delays in completing the first rating, it may behoove researchers and site managers to review these procedures with raters every few months to minimize loss of skill.

Clinical expertise was also considered as a possible confounding variable. To make a valid comparison with the two outlier sites with the worst inter-rater reliability, the two sites with the best inter-rater reliability were chosen as comparisons. This possible confounding variable did not appear to create any significant variation in reliability, as the two sites with the best and worst reliability had nearly the same percentage of expert clinicians as assessed by specialty certification (80% in the sites with the best reliability compared to 75% in the sites with the worst reliability). Furthermore, a Mann-Whitney U test was performed to statistically compare the sites with best and worst reliability and it yielded no significant differences ($p=0.62$).

www.manaraa.com

While it has been argued that specialist certification is an accurate demonstration of expertise,[130] others have contended that the metric of specialty certification is not the best metric for determining expertise.[131] Another method that has been utilized to distinguish between expert and novice clinicians has been years of experience.[111-113] To minimize the risk of misunderstanding the level of expertise between the sites with best and worst reliability, a further comparison utilizing years of experience as an alternative metric for expertise was performed. A Mann-Whitney-U test resulted in p=0.44. Thus, the two sites with the best and worst reliability demonstrated no significant differences in years of experience of the raters or in specialist certification, and those cannot be considered viable factors in the differences between those sites with better or worse reliability of rating.

Thus, as the training for this study was purposefully generalizable (online narrated lecture (https://youtu.be/a-QiJ5-bKKQ)[107] combined with the assigned reading of a published, peer-reviewed paper[2]), this demonstrates that the reliability of this rating system is not contingent upon experience or expertise.

Aim 2: Compare level of functional limitation between irritability groups

Function related to shoulder symptom irritability

In discussing the concept of shoulder symptom irritability with many physical therapists at conferences across the country, the most common misperception has been that the term shoulder symptom irritability is synonymous with pain level. Shoulder symptom irritability is the tissue's readiness to accept physical stress and theoretically relates to the tissue's physical status and the degree of inflammatory activity present. While many of the proposed components of shoulder symptom irritability are varying

www.manaraa.com

constructs of pain with and without movement, one component involves the assessment of functional limitation.[2,14] Thus, we assessed if functional limitation is a necessary component of the SSIC system.

Of the 101 subjects rated for shoulder symptom irritability in this study, raters matched classification level for 68 of the subjects. As previously described, to preserve the validity of this analysis, only those subjects receiving the same classification by both raters were included. One subject did not complete the FOTO FS and thus there are only 67 subjects included in the analysis of FOTO FS.

Statistically significant differences were found for all three patient-reported functional outcome measures (PSS, ASES, and FOTO FS). Furthermore, Bonferroni post hoc testing demonstrated significant differences between all shoulder symptom irritability groups for the PSS ($p<0.001$), ASES ($p<0.007$), and FOTO FS ($p<0.003$). These results demonstrate that level of functional limitation is lower in patients with low shoulder symptom irritability, moderate in subjects with moderate shoulder symptom irritability, and higher in subjects with high shoulder symptom irritability and may help inform the classification of shoulder symptom irritability.

Therefore, since the level of functional limitation differentiate well (ES = 3.20-6.80) between shoulder symptom irritability classification levels, further testing is warranted to determine the level of influence it has on predicting shoulder symptom irritability. Additionally, due to the strong ability of functional limitation to differentiate between shoulder symptom irritability classification levels, further exploratory analyses to determine cut-off scores for the three patient-reported functional outcome measures were performed.

www.manaraa.com

Patient-reported functional outcome measure cut-off scores

Receiver operating characteristic (ROC) curves were analyzed for each of the three patient-reported functional outcome measures to determine preliminary cut-off scores. As mentioned earlier in comparison to a pilot study by Kareha and colleagues, one possible reason for the lower reliability in this study when compared to pilot data was the vague description of low, moderate or high functional limitation scores in this study compared to the specific cut-offs provided during the pilot study. While this was convenient for the pilot study, there is no data to drive these cut-off scores and thus they were removed in the development of this study to more accurately represent the state of the evidence.[2] Thus, the development of cut-off scores for the three patient-reported functional limitation outcome scores may aid clinicians in accurate classification of shoulder pain irritability.

The cut-off scores derived in this study demonstrate moderate to excellent likelihood ratios (3.84-11.56) for determining high shoulder symptom irritability and small to excellent likelihood ratios (2.66-10.26) for determining low shoulder symptom irritability.[128] Even in the larger group which included all 202 ratings, including 66 SSIC ratings that did not match, the agreement between these cut-off scores and the SSIC level was still good (Table 13). In this secondary analysis, the cut-off scores with the best agreement were the PSS Function Subscale (69%), PSS Total Score (68%), ASES Total Score (66%), and PSS Pain Subscale (62%).

These results demonstrate that the best patient-reported outcome measure to help determine shoulder symptom irritability is the PSS as it has the greatest overall agreement with rater SSIC selection. Additionally, the PSS subscales of pain and

www.manaraa.com

function have a greater correlation with rater SSIC selection than the ASES pain and function subscales and FOTO functional score. This is interesting in light of a recent decision of the American Academy of Orthopaedic Surgeons to recommend the use of the ASES or Oxford Shoulder Score as core patient-reported outcome measures in all future research.[137] A possible explanation for this seeming inconsistency is that the ASES is better for surgical decision-making while the PSS is better for non-surgical decision making, but this hypothesis would require additional research. Also, while the development of these cut-off scores for the patient-reported outcome measures is promising, further work is needed to validate these results in other samples prior to clinical use.[128]

Additionally, these results are interesting because lower functional ability appears to be very indicative of high shoulder symptom irritability, whereas higher functional ability is not as strongly indicative of low shoulder symptom irritability. While no research exists that can currently explain these results, it is logical that those patients with high shoulder symptom irritability would have very low functional ability, while the difference between moderate and low irritability may be more dependent upon personal functional needs and desires.

For example, if a patient presented with high levels of pain and difficulty reaching to shoulder level, most patients of all ages and ability levels would likely consider that level of functional limitation to be high. However, if a patient presented with difficulty with lifting 10 pounds overhead, a 25-year-old construction worker might consider that moderate functional limitation, but a 95-year-old sedentary person might consider that a very minimal functional limitation. Thus, while the differentiation of function at the

www.manaraa.com

higher end is likely consistent with the construct of activity limitation as a component of the International Classification of Functioning, Disability, and Health (ICF) model, the differentiation of functional limitation levels at the lower end may be more closely related to the construct of participation in the ICF model.[138]

Future research should test this new hypothesis that patients with high shoulder symptom irritability would have very low functional ability, whereas the difference between moderate and low irritability may be more dependent upon personal functional needs and desires. If the hypothesis is supported, then it is likely that functional limitation is most helpful in dichotomizing high shoulder symptom irritability from all other levels of shoulder symptom irritability. Whereas, understanding the patient's perspective regarding participation desires, environmental factors, and personal factors may be a beneficial addition to more effectively determine those patients who would benefit from interventions with moderate to high tissue stress. Thus, the addition of a simple component to measure participation restriction, environmental factors, or personal factors may be necessary to further improve the usefulness of the Shoulder Symptom Irritability Classification system and thus aid in the appropriate dosage of non-surgical intervention.

Pain related to shoulder symptom irritability

The SSIC has an overwhelming predominance of pain-related components as it is essentially investigating the construct of how pain interacts with function. Thus, additional analyses were performed beyond the aims of the study to obtain a better understanding of the data.

www.manaraa.com

Other studies have demonstrated that to establish an accurate pain rating, one must ask the patient multiple questions rather than just a single question regarding pain.[139] Thus, it was expected that the PSS Pain Subscale, which is a composite of 3 items, would have a greater relationship with SSIC groups than the ASES Pain Subscale. Ordinal regression was utilized to compare the differences in influence the two subscales have upon the classification of shoulder symptom irritability. In this analysis, the PSS Pain Subscale did significantly influence the classification of shoulder symptom irritability, but the ASES Pain Subscale did not. These results may indicate the PSS Pain Subscale is a better scale than the ASES Pain Subscale for informing appropriate rehabilitation diagnosis.

Additionally, since clinicians are very busy and frequently do not feel they have time to ask multiple pain questions, it would be useful to understand the influence of specific items within the PSS Pain Subscale to more efficiently make treatment decisions. Since 2 of the PSS Pain Subscale items involve how pain influences function, "pain with normal activities (eating, dressing, bathing)" and "pain with strenuous activities (reaching, lifting, pushing, pulling, throwing)", we anticipated that they would be much more influential upon the classification of shoulder symptom irritability than the remaining item, "pain at rest with your arm by your side." This hypothesis was supported as both the activity-based pain items significantly influenced the classification of shoulder symptom irritability, while the resting pain item did not significantly influence classification. These results aid in developing a better understanding of the construct validity of the SSIC system involving how pain influences function.

www.manaraa.com

Pain and function on shoulder symptom irritability

The parameter estimates (-0.12 for the PSS function subscale and -0.44 for the PSS pain subscale) were small, demonstrating that to make a change in irritability level, one must have a multiple point change in pain or functional limitation scores. This makes sense logically as scores from both the pain subscale and function subscale of the PSS result in a much wider range than the 3-point ordinal scale of shoulder symptom irritability.

The parameter estimates of both the pain and function subscales of the PSS demonstrated a significant impact on shoulder symptom irritability (PSS Pain subscale $p<0.001$, PSS Function subscale $p=0.006$). However, the pain subscale demonstrated significantly more impact on shoulder symptom irritability as the 95% confidence intervals do not overlap and the Wald statistic is much greater for the PSS Pain subscale than it is for the PSS Function subscale.

Furthermore, this statistical finding was reinforced visually by scatter plots of the individual subscales and shoulder symptom irritability level (Figures 25-26). These figures show a much tighter cluster of PSS Pain scores within shoulder symptom irritability levels (Figure 25) than with the PSS Function scores (Figure 26). This observation demonstrates that PSS Pain scores more closely follow SSIC selection than PSS Function scores.

Aim 3: Determine if the level of irritability logically guides the chosen intervention

The point of classifying shoulder symptom irritability is to improve clinical decision-making for the selection and intensity of intervention. Studies have demonstrated that appropriately matching treatment strategy to diagnostic classification

www.manaraa.com

results in improved outcomes in patients with neck pain and low back pain.[29,30,66] To begin to establish a better understanding of whether SSIC dictates treatment strategy, the correlation between rater selection of SSIC group and rater selection of treatment strategy was analyzed.

To minimize threats to internal validity from educational factors (such as biasing) and social factors (such as imitation), the rater training specifically did not emphasize what treatment strategy or intensity should be prescribed beyond the theoretical framework of why this study is important. Rather, the emphasis was placed on the classification process, in order to accurately determine the ability of raters to be trained to reliably classify shoulder symptom irritability. Therefore, while we cannot completely eliminate the chance that clinical decisions were dictated by previously published reports, our rater training methods reduce that risk to the greatest degree possible while still establishing an adequate and generalizable training method for appropriate levels of reliability.

Each of the 101 subjects was rated by two separate raters, resulting in 202 clinical decisions regarding overall treatment strategy. Despite some dissent to the idea that those providers with less experience and expertise would rate subjects with lower reliability,[140] there has been evidence demonstrating differences in reliability among raters with varying levels of experience and expertise.[123,124,126] Therefore, due to the high variability of experience and expertise of raters in our sample, expected levels of correlation were moderate.[126]

However, based on our results, rater agreement was excellent with PABAK-OS = 0.82 and 80% agreement despite these concerns. While these results do not establish

www.manaraa.com

final construct validity, they do provide evidence that the classification of shoulder symptom irritability may impact the choice of treatment strategy prescribed to the patient. Therefore, further research is indicated to determine if following treatment strategies matched to SSIC results in improved patient-centered outcomes.

Further analysis was performed to determine if there was a difference in agreement between specialist providers and non-specialist providers. While there were no statistical differences between specialists and non-specialists, a trend was observed in which matched intervention was more likely to be selected by non-specialists than specialists. This was intriguing, as the expectation was that specialist providers would be more likely to select a matched treatment strategy. As this trend was not statistically significant it is most likely that the trend was errant.

Given that there was a trend, but no significant differences regarding the a priori selected measure of expertise (specialist certification) additional metrics were investigated to determine if there were any factors likely to improve or diminish the matching of treatment strategy to SSIC. Other metrics of expertise that have been utilized in the literature have been years of experience,[111-113] but there is no specified number of years that signify expertise. Thus, the data were analyzed by separating the raters into those practicing 0 to 5 years and greater than 5 years consistent with the "New Professionals" designation by the APTA.[114] A non-statistically significant trend was found that seemed to indicate those practicing for more than 5 years may be better at choosing a treatment strategy that matches the SSIC.

In an effort to follow this theory, the same analysis was performed after dichotomizing the raters into those practicing 0 to 10 years and greater than 10 years as it

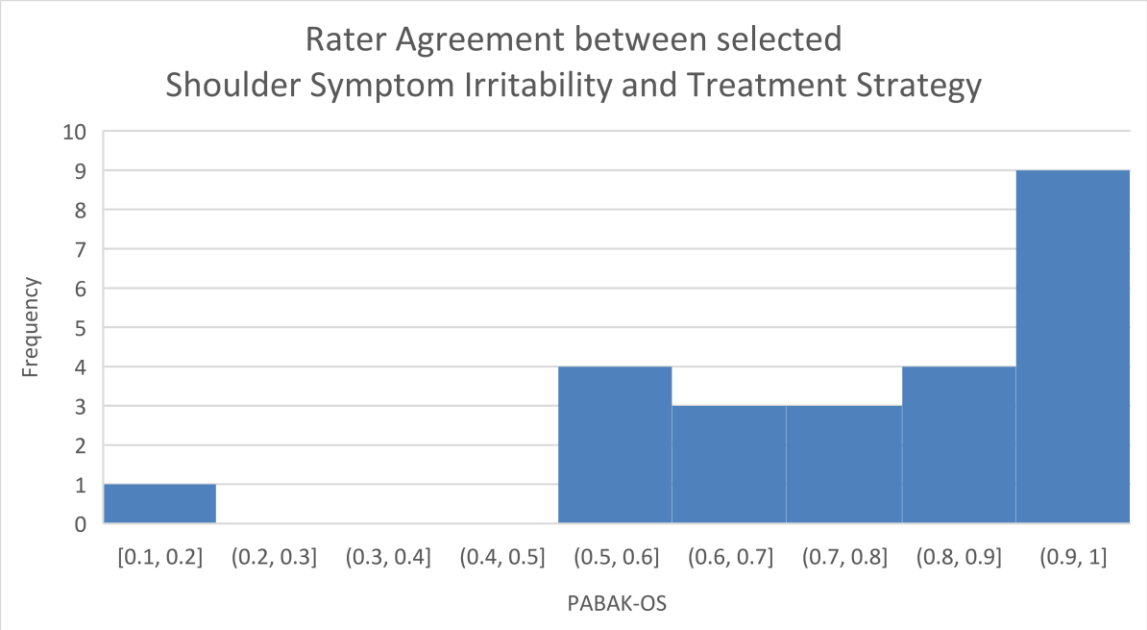
www.manaraa.com

would be expected that those practicing longer than 10 years would be as good if not better than those practicing for fewer years. It is noteworthy that while there was not a large enough sample for appropriate power to perform a non-inferiority study, we were only looking for trends and thus the power limitation was acceptable. However, the trend did not continue, and in fact, the trend did not replicate itself at all in the 10-year group.

It is also logical to consider that possibly a few raters inappropriately skewed the matching of treatment strategy toward the positive or toward the negative. The plot of agreement of the raters between the selected SSIC and the treatment strategy demonstrated a negative skew (Figure 27), but even with the exclusion of the outliers and also those with the highest data contribution, no significant changes in agreement were realized. These results reinforce the construct validity of the SSIC that it does, in fact, influence the selection of treatment strategy. However, the true test of construct validity is not simply that the SSIC influences the selection of treatment strategy, but that the treatment strategy selected via the SSIC enhances patient-centered outcomes.[2]

www.manaraa.com

Figure 27: Rater Agreement between selected Shoulder Symptom Irritability and Treatment Strategy



Post-hoc power analysis for the analyses in the third aim of this study demonstrated power (1-β) of only 19%. Therefore, if the trends found above were not errant, demonstrating type II error, the noted trend could be a result of the specialists incorporating another construct that is not currently included in the SSIC system. Recent studies have found that exposure to psychosocial factors at work increases the odds of reduced functional level due to neck and shoulder symptoms.[141,142] Additionally, psychosocial factors have demonstrated predictive validity for persistent shoulder pain after breast cancer surgery[143] and overall functional outcome following an episode of care involving physical therapist services.[144] Furthermore, the addition of psychosocial factors would be consistent with the ICF constructs of participation restriction, environmental factors, and personal factors that may have influenced the decreased ability of functional limitation (activity limitation in ICF language) to differentiate

www.manaraa.com

between low and moderate SSIC groups, when compared to the ability of functional limitation to differentiate high and moderate SSIC groups.

While no studies to date, to the author's knowledge, have investigated if altering treatment based on psychosocial factors for shoulder pain improves patient-centered outcomes, this concept has been explored in spinal pain. Researchers have found that the expectation of successful outcome is the greatest predictor of success for cervical manipulation for patients with neck pain.[145] Additionally, high fear-avoidance has been found to decrease the likelihood of successful outcome following lumbar manipulation for patients with low back pain.[146] Experts have also suggested that patient expectation of successful outcome is likely to have a strong role in the outcome of care for musculoskeletal pain.[147] Since no research has specifically investigated the effects on modulating intervention strategy based upon psychosocial factors in patients with shoulder pain, future work should consider psychosocial factors either as an additional variable to supplement the SSIC or as an added factor to the five current components of the SSIC system.

Implications

This study was designed to provide a better understanding of the reliability and validity of the Shoulder Symptom Irritability Classification System (SSIC), a previously untested component of the Staged Approach to Rehabilitation Classification: Shoulder Disorders (STAR-Shoulder) diagnostic classification system. The results of this study demonstrate the Shoulder Symptom Irritability Classification System has satisfactory inter-rater reliability for use in clinical practice that is comparable to other widely accepted and utilized classification systems.[16,24,42,66,121-126] Furthermore, it does not

www.manaraa.com

appear that experience with utilizing the SSIC nor expertise in practice affects the reliability of the classification system; therefore, it can be taught to entry-level clinicians to improve communication without concern that they will be less reliable in classification than their experienced peers. The training for this classification system involves the reading of a freely accessible, peer-reviewed paper (https://academic.oup.com/ptj/article/95/5/791/2686487)[2] and a freely accessible online tutorial (https://youtu.be/a-QiJ5-bKKQ)[107] making this system highly generalizable to physical therapists. Therefore, the results of this study should serve as a foundation for future work for refinement as a component of the STAR-Shoulder diagnostic classification system.

These data also provide evidence of the importance of functional limitation status in the SSIC system, especially when differentiating between high shoulder symptom irritability and moderate shoulder symptom irritability. Furthermore, it provides evidence that the Penn Shoulder Scale may be more influential, with regard to pain and function subscales, than other patient-reported outcome measures in influencing the classification of shoulder symptom irritability. Additionally, these results aid in developing a better understanding of the construct validity of the SSIC system involving how pain influences function.

This study also provides evidence of the impact function and pain constructs both have upon the resulting Shoulder Symptom Irritability Classification. One of the greatest limitations in previous studies,[75] clinical guidelines,[27] and commentaries[2,14] was there were no cut-off scores for the patient-reported functional outcome measures and whatever cut-offs were utilized, were arbitrary. This study provided the first derivation

www.manaraa.com

of cut-off scores to help inform judgment about irritability. As these cut-off scores have not yet been validated in other samples, further work is needed prior to being utilized in clinical practice. Additionally, these results question whether there may be an additional component to shoulder symptom irritability related to participation restriction, environmental factors, or personal factors to improve the reliability and validity of the SSIC system, and ultimately the STAR-Shoulder diagnostic system to adequately inform practice patterns.

Lastly, this study begins to build the necessary framework of correlation between diagnostic classification and treatment decision-making. Due to the excellent correlation between shoulder symptom irritability classification and intended treatment strategy, the initial analysis of construct validity is promising. However, before this classification system is fully utilized in clinical practice, further research is necessary to determine if the treatment strategy selected via the SSIC enhances patient-centered outcomes.

It is anticipated that the SSIC system will be integrated with health condition (pathoanatomy) and body functions and structure (impairments) as recommended in the STAR-Shoulder diagnostic classification system to appropriately prescribe rehabilitation intervention and reduce unwarranted variation in clinical practice.[2] Ultimately, the reduction in unwarranted variation is expected to result in reduced costs for the health care system and improved functional outcomes for patients.

Recommendations

While this study has demonstrated good inter-rater reliability of the shoulder symptom irritability classification system, future studies should aim to validate the inter-rater reliability of this study utilizing sites and raters from multiple regions of the country

www.manaraa.com

and obtaining a sample of raters that are more representative of the nation's proportion of doctorally-trained and board-certified physical therapists. Additionally, it would be prudent to attempt to analyze rater pairs of expert clinicians only and novice clinicians only to determine if there are any differences in inter-rater reliability within groups. However, given the inter-rater reliability of the SSIC compared to other widely accepted and utilized classification systems, it is satisfactory for reliable use in clinical practice.[16,24,42,66,121-126]

To improve the blinding of the inter-rater reliability component but not compromise the live patient examination model, future studies could include single day, serial patient examinations of persons with shoulder pain outside of normal clinical practice. Study personnel would need to be present for the entire time to restrict communication between raters. This manner would be expensive but would assure blinding is maintained.

The general purpose of the SSIC is to improve patient care and greater emotional intelligence has been correlated with improved clinical outcomes.[132] Therefore, it is recommended that future studies include measures of emotional intelligence of the raters to aid in the analysis of what factors improve the reliability of SSIC selection and subsequent selection of treatment strategy. Furthermore, based upon the diminished ability of functional limitation to influence the differentiation between low and moderate SSIC groups, patient-centered measures of participation restriction, environmental factors, and personal factors should be included to determine if there are any additional factors that may influence the selection of SSIC and the subsequent selection of treatment strategy.

www.manaraa.com

Additional research will be necessary to determine the validity of the cut-off scores that were derived in this study. Likely this could be a secondary aim of a subsequent study if it is powered well enough; and due to the high effect size differentiating between SSIC levels, this recommendation is a realistic expectation of future studies. Also, once these cut-offs have been validated, it will be important to decipher if there continues to be one specific patient-reported outcome measure that is more helpful in determining non-surgical intervention strategy. If one can be found, our recommendation is that the most helpful patient-reported outcome measure for non-surgical care be utilized nationwide in future research and clinical practice involving shoulder pain, in order to improve patient care and facilitate better communication and comparison across samples.

Further analysis is needed to determine if treatment matched to the patient's shoulder symptom irritability results in improved functional outcome and/or fewer visits to reach the patient's therapeutic goals. Likely, this would be best evaluated in a pragmatic trial in which shoulder symptom irritability is classified and then specific interventions performed were classified into low intensity, moderate intensity or high intensity at each visit. Raters would be evaluated for emotional intelligence and in addition to the measures utilized in this study, additional measures for participation restriction, environmental factors, and personal factors would need to be collected to determine if the inclusion of one or more of these factors can aid in improving inter-rater reliability and effectiveness of care. Each selected intervention would be classified a priori into low intensity, moderate intensity, and high-intensity interventions. This study would likely be costly due to the increased burden on the rater and increased risk of

www.manaraa.com

attrition of subjects. The projected sample size of this type of study would likely need to be two to three times the size of the present study, based upon our regional hospital network's unpublished completion rate.[148]

After the episode of care was completed, patients would be grouped into those who received matched treatment strategies and non-matched treatment strategies, based upon a majority threshold of treatments that matched intensity of intervention to SSIC. Functional, satisfaction, and pain outcomes would be compared between groups. Data would be analyzed via independent t-test to determine statistical differences between groups.

Limitations and Delimitations

Limitations

A limitation of this study is the use of physical therapists and consecutive patients from a single regional hospital network. However, the regional hospital network encompasses over 40 locations with over 90 physical therapists, 28 of which were clinical specialists. In this manner, we attempted to obtain a wide sample from across the regional hospital network. And while not feasible for the present study due to funding, future studies could be improved by utilizing multiple sites from multiple regions across the country. It would also be prudent to analyze rater pairs of expert clinicians only and novice clinicians only, as the lack of ability to do so in the present study was another limitation.

Another limitation of this study was the significant difference in the proportion of doctorally-trained physical therapists and board-certified specialists in our regional hospital network when compared to the much smaller number in both categories

www.manaraa.com

nationally. And, while the results of this study do not appear to support any differences based upon these criteria, there may be differences discovered in samples with more representative populations of the educational level of the nation's physical therapists.

While we tried to control for blinding and social bias with specific instruction, there was, of course, no possible way given the level of funding and the present methodology, to guarantee blinding throughout the study. However, the analyses do appear to refute the idea that repeated rating developed any improvement in inter-rater reliability, thus diminishing the likelihood that blinding had failed.

Another limitation to consider is that we do not know the effect of the rater training. It is possible that the SSIC is reliable without training. It is also possible that the training provided, while generalizable would be better served in a live, synchronous classroom setting.

Although the FOTO functional status instrument requires permission to access and use, FOTO, Inc. offers access to their data at no cost to researchers. Additionally, it is easy to administer, score, and has a low patient burden.[149] However, the FOTO functional status instrument is a proprietary measure, and while it is utilized nationally and internationally, it is not likely that it will be universally used due to its proprietary nature. Thus, despite the concern of increased patient burden, two widely utilized, free, patient-reported outcome measures were utilized (the PSS and ASES) to improve generalizability.

Delimitations

A delimitation of this study would be the presumption that the criteria for expert clinicians are met by the requirements for attaining certification by the American Board

www.manaraa.com

of Physical Therapy Specialties. While this has been argued in previous studies,[130,131] it appears to be the best proxy for expertise besides a patient-outcome based model[131,150] which was unrealistic for a study of this scope and without significantly greater funding.

Due to the inter-rater reliability nature of this study, the need to have more than one clinician trained and available to rate the subjects at all times likely lead to some potential subjects not being asked to participate. It could have been possible to perform this study via video analysis, but as the lack of live evaluation was a limitation of former studies,[66] it was decided that the benefits of live evaluation outweighed the limitations of having more than one clinician trained and available to rate the subjects at all times at each site.

Additionally, the data in Aim 2 was limited to only those subjects who had complete agreement between raters. This reduced the power of this part of the study, but the risk to internal validity by utilizing the subjects without complete agreement was a greater threat to the study than the limitation of power to aim 2. Even then, the post-hoc power (1-β) analysis of ≥ 87% for all scales except for the ASES Pain subscale, for which power (1-β) was 78%, was excellent and thus demonstrates that the study was powered appropriately for all three aims.

Finally, the data obtained regarding intended treatment strategy is simply that, intended. While longitudinal outcomes data would be ideal, the aim of utilizing intended treatment strategy was to determine if further investigation utilizing longitudinal outcomes data is necessary, given the time and financial implications of such a study.

www.manaraa.com

Summary

Background

Diagnosis has been integral in western medicine and is aimed at guiding the treatment approach, determining a prognosis, and succinctly communicating the signs and symptoms of the patient to other providers to aid in the patient's recovery.[47] For a diagnosis to be meaningful, it is implicit that a diagnosis should direct the most appropriate intervention for that condition, determine a prognosis, and that diagnoses should be mutually exclusive from one another.

Clinicians regularly make decisions regarding intervention intensity based upon diagnosis, but evidence has demonstrated that expert clinicians to do not utilize pathoanatomical diagnosis to make these decisions.[6] This may be due to a lack of correlation of pathology to activity limitations, participation restrictions, and symptoms. Therefore, a need exists to develop an adequate diagnostic system beyond the single classification construct of the anatomic structure implicated, in order to more accurately guide treatment decision making and inform prognosis.[2]

Treatment-based diagnostic processes have been proposed for the non-operative management of low back pain, neck pain and due to the heterogeneity of patient presentation and resultant poor outcomes of interventional studies within those diagnostic categories.[8,62,63,65] While this is also true regarding shoulder disorders, the literature on the prognosis of shoulder disorders does demonstrate a correlation between pathoanatomic diagnostic classification and prognosis.[27,59,67-70] Thus, an optimal classification system to improve treatment decision-making for patients with shoulder disorders would encompass pathoanatomy, shoulder symptom irritability, and physical

www.manaraa.com

impairments giving birth to the Staged Approach for Rehabilitation Classification: Shoulder Disorders (STAR-Shoulder) diagnostic system.[2]

Symptom irritability has been utilized extensively by clinicians and researchers for many years.[9,15,16,61,63,72,73] It is important to note that symptom irritability is not synonymous with the acuity of symptoms or simply pain level.[8] Multiple experts in physical therapy have proposed criteria for symptom irritability from which to base clinical decisions for intensity.[7-9,13,14] However, the reliability of these classifications have been tested and found to be poor to moderate, and none have been specific to the shoulder.[13,16] Kelley and McClure[14,27] proposed a method of classifying symptom irritability specifically for the shoulder, but to our knowledge, no studies have determined the reliability and validity of Shoulder Symptom Irritability Classification (SSIC).

Purpose

The objective of this dissertation is to begin to establish the reliability and construct validity of shoulder symptom irritability as one part of the STAR-Shoulder classification system to guide refinements.

Design

This study employed a prospective quasi-experimental observational design utilizing single-blinded repeated measures (specific aim 1), followed by cross-sectional analysis (specific aims 2 and 3). (Figure 1)

Methods

Ethics approval was obtained from the Institutional Review Boards of St. Luke's University Health Network (2016-61) and Nova Southeastern University (2016-379).

www.manaraa.com

Raters were recruited via email and personal request, consented in person and demographic data on the raters was collected. The raters were then trained utilizing a freely accessible, peer-reviewed paper[2] and a freely accessible online tutorial.[107] Patient subjects were recruited from a convenience sample of consecutive patients presenting for physical therapy consultation for shoulder pain, not extending to the neck. Raters recorded the shoulder symptom irritability level and selected the appropriate treatment strategy (intensity) for each of the subjects.

Data Analysis

Descriptive statistics were used to characterize both raters and patients. Prevalence-adjusted, bias-adjusted Kappa for ordinal scales (PABAK-OS) and observed agreement were the primary measures of inter-rater reliability and reliability of matched treatment strategy selection. For evaluation of statistical significance, a two-tailed confidence interval was utilized with α set to 0.05, and the null hypothesis was that the PABAK-OS is <0.40.[84] Analysis of variance with post-hoc analysis was used to compare functional disability across different levels of irritability. Receiver operating characteristic curve analyses for the different patient-reported functional outcome scales and subscales were used to determine the cut-off values that would maximize the sensitivity and specificity of each scale. Lastly, ordinal regression was utilized to compare the strength of patient-reported pain and disability in the determination of shoulder symptom irritability.

Results

101 consecutive subjects with primary complaints of shoulder pain were assessed by pairs of blinded raters (24 raters in total). Of the 24 raters that submitted patient data

www.manaraa.com

for this study, the mean age was 33.9 (+/-7.3) years with a mean of 8.1 (+/-6.7) years of experience in clinical practice (Table 3). The mean age of the 101 consecutive patients included in the study analyses was 56.0 (+/-16.0) years, with females accounting for 65.3% of the sample (Table 4).

All 101 subjects were included in the inter-rater reliability analysis. The inter-rater reliability of the SSIC system was PABAK-OS = 0.69 (95% Confidence Interval [CI] = 0.59-0.78) and the percent agreement between raters was 68% (Table 5). There was no significant improvement or degradation of reliability between groups that have had increased experience rating subjects (Tables 8-9) or more expertise in practice (Figure 6) when compared to those groups that have had less experience or less expertise.

Analysis of variance demonstrated significant differences in functional limitation between irritability groups for the PSS, ASES, and FOTO FS ($p<0.001$) (Table 10) and revealed large effect sizes of patient-reported outcome scores on shoulder symptom irritability group ranging from 3.20-6.80 (Table 11).

Receiver operating characteristic curve analyses for the patient-reported functional outcome scales and subscales were used to determine the cut-off values that would maximize the sensitivity and specificity of each scale (Figures 9-21). The cut-off scores were then compared to all rater choices and found the scales with the best agreement were the PSS Function Subscale (69%) and the PSS Total Score (68%) (Table 13).

Ordinal regression was utilized to compare the two pain subscales from the PSS and ASES. As summarized in Table 14, the PSS Pain subscale significantly influenced the selection of SSIC, while the ASES Pain Subscale did not. Additionally, the two items

www.manaraa.com

of the PSS Pain Subscale that involve how pain influences function also influenced the selection of SSIC (Table 15).

Ordinal regression was also utilized to establish the degree of influence pain has upon SSIC compared to function. Parameter estimates were -0.44 (95% CI -0.66, -0.22) for the pain subscale and -0.12 (95% CI -0.21, -0.04) for the function subscale (Table 16) demonstrating the superior influence of the pain subscale on shoulder symptom irritability when compared to the function subscale.

As shoulder symptom irritability is designed to determine treatment strategy and intensity, rater agreement between SSIC and treatment strategy was found to be PABAK-OS = 0.82 (95% CI 0.75, 0.88) with 80% agreement (Table 17). Thus, the hypothesis that the level of shoulder symptom irritability is moderately correlated (K>0.40 and agreement>50%) with planned intervention intensity is accepted. No significant differences in agreement between shoulder symptom irritability and treatment strategy were found when data was dichotomized between specialist and non-specialist groups (p=0.56). Additionally, years of experience did not significantly alter the matching of treatment strategy to shoulder symptom irritability classification (Tables 20-23), nor did any outliers significantly influence the final results (Tables 25-26).

Discussion

The patients in the present study were of comparable demographics to samples found in epidemiological studies improving the generalizability of the results.[117-119] Also, the factors in which raters did not resemble the national population of physical therapists[115,116] did not seem to influence any aspects of the study.

www.manaraa.com

Inter-rater reliability

While the hypothesis that inter-rater reliability of the SSIC has a PABAK-OS of >0.60 cannot be supported with 95% confidence, other well accepted and commonly utilized scales have demonstrated similar or poorer inter-rater reliability.[16,24,42,66,121-126] Additionally, no individual sites or degree of experience or expertise appeared to significantly influence the reliability of the SSIC. This demonstrates that the inter-rater reliability of the SSIC system is not contingent upon experience or expertise; and that it is similar, if not better than, many widely accepted classification systems[16,24,42,66,121-126] and can be considered sufficiently reliable for clinical use.[24]

However, due to the wide variability of individual site reliability and the lack of difference due to experience or expertise, other factors must be considered. Since social awareness is one of the four components of emotional intelligence (EI),[133] it is logical that EI may affect the reliability of the SSIC as observational analysis is a major component of accurately classifying shoulder symptom irritability.

Component Analysis

Function related to shoulder symptom irritability

These results demonstrate that level of functional limitation is lower in patients with low shoulder symptom irritability, moderate in subjects with moderate shoulder symptom irritability, and higher in subjects with high shoulder symptom irritability and may help inform the classification of shoulder symptom irritability. Since the level of functional limitation has a strong effect (ES = 3.20-6.80) on shoulder symptom irritability, further testing is warranted to determine the level of influence it has in predicting shoulder symptom irritability.

www.manaraa.com

Patient-reported outcome measure cut-off scores

The cut-off scores derived via ROC curve analysis in this study demonstrate moderate to excellent likelihood ratios (3.84-11.56) for determining high shoulder symptom irritability and small to excellent likelihood ratios (2.66-10.26) for determining low shoulder symptom irritability.[128] These results demonstrate that the best patient-reported outcome measure to help determine shoulder symptom irritability is the PSS, because it has the greatest overall agreement with rater SSIC selection. An important observation is that lower functional ability appears to be more indicative of high shoulder symptom irritability than higher functional ability does in indicating low shoulder symptom irritability. It is logical that those patients with high shoulder symptom irritability would have very low functional ability, whereas the difference between moderate and low irritability may be more dependent upon personal functional needs and desires. This is consistent with the participation restriction, environmental factors, and personal factors aspects of the International Classification of Functioning, Disability, and Health (ICF) model.

Pain related to shoulder symptom irritability

We anticipated that since two of the PSS Pain Subscale items involve how pain influences function, they would be much more influential upon the classification of shoulder symptom irritability than the remaining item of pain at rest. This hypothesis was supported as both of the activity-based pain items significantly influenced the classification of shoulder symptom irritability, while the resting pain item did not significantly influence classification. These results aid in developing a better

www.manaraa.com

understanding of the construct validity of the SSIC system involving how pain influences function

Pain and function on shoulder symptom irritability

Due to the overwhelming predominance of pain-related components of shoulder symptom irritability, it was important to decipher if pain subscales demonstrated stronger prediction of shoulder symptom irritability levels than functional limitation subscales. Ordinal regression found that both pain and function significantly influence SSIC, but the pain subscale demonstrated significantly more impact on shoulder symptom irritability.

Impact on treatment strategy selection

The purpose of classifying shoulder symptom irritability is to improve clinical decision making for the selection and intensity of intervention. To begin to establish a better understanding of whether SSIC dictates treatment strategy, the relationship between rater selection of SSIC group and rater selection of treatment strategy was analyzed.

In this study, the relationship between rater selected SSIC and treatment strategy was excellent. Additionally, experience and expertise did not significantly influence the matching of SSIC to the selection of treatment strategy. While these results do not establish final construct validity, they do provide evidence that the classification of shoulder symptom irritability may impact the choice of treatment strategy prescribed to the patient. Therefore, further research is indicated to determine if following treatment strategies matched to SSIC results in improved patient-centered outcomes.

www.manaraa.com

Recommendations

The results of this study should serve as a foundation for future work for refinement as a component of the STAR-Shoulder diagnostic classification system. This future refinement should include patient-centered measures of participation restriction, environmental factors, and personal factors to determine if there are any additional influences that may impact the selection of SSIC and the subsequent selection of treatment strategy. Also, as the patient-reported outcome measure cut-off scores derived in this study have not yet been validated in other samples, further work is needed prior to being utilized in clinical practice. Finally, before this classification system is fully implemented in clinical practice, further research is necessary to determine if the treatment strategy selected via the SSIC enhances patient-centered outcomes.

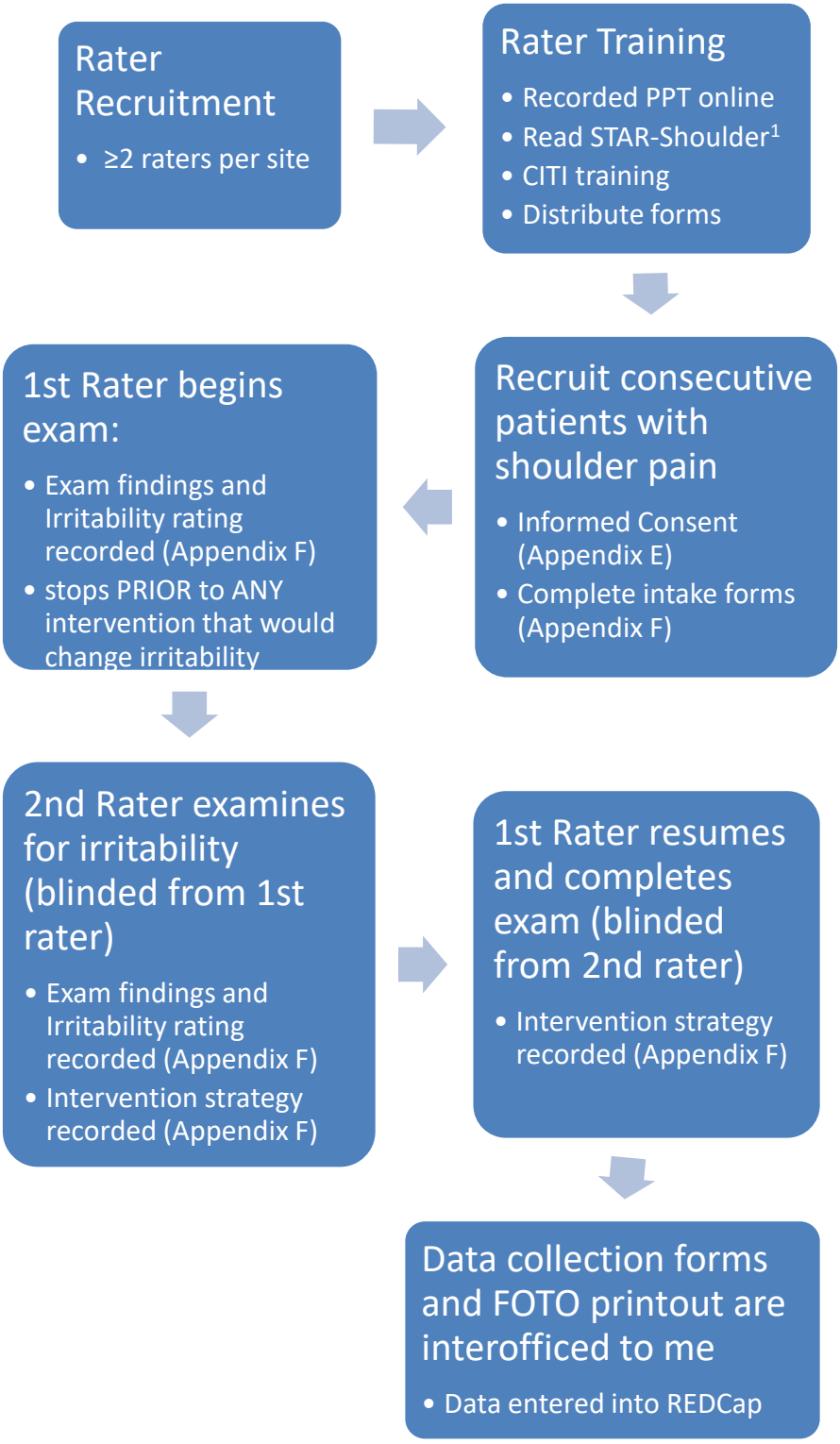
Clinical Significance

The shoulder symptom irritability classification scale is reliable and clinically useful for improvement of communication between medical providers. It also has the potential to improve patient outcomes by directing the most efficient use of resources with the appropriate dosage. It is anticipated that the SSIC system will be integrated with health condition (pathoanatomy) and body functions and structure (impairments) as recommended in the STAR-Shoulder diagnostic classification system to appropriately prescribe rehabilitation intervention and reduce unwarranted variation in clinical practice.[2] Ultimately, the reduction in unwarranted variation is expected to result in reduced costs for the health care system and improved functional outcomes for patients.

www.manaraa.com

APPENDICES

Appendix A: Recruitment & Logistics



www.manaraa.com

Appendix B: Rater Information

Rater Information

Name: _____

Age: _____

Years of Practice: _____

Advanced Certifications Held:

- ☐ OCS
- ☐ SCS
- ☐ FAAOMPT
- ☐ Other _____

Gender: ☐ Male ☐ Female

Entry Level Degree:

- ☐ BS
- ☐ MS
- ☐ DPT

Highest Earned Degree:

- ☐ BS
- ☐ MS
- ☐ DPT
- ☐ PhD, ScD, EdD

www.manaraa.com

Appendix C: Inclusion/Exclusion Criteria

For Office Use Only:
Subject #_____

Inclusion Criteria

- ☐ 18 years old or older
- ☐ Chief complaint of shoulder pain

Exclusion Criteria

- ☐ Pain or symptoms distal to the elbow
- ☐ History of ipsilateral shoulder surgery
- ☐ Active or passive cervical spine ROM reproduces shoulder pain
- ☐ Positive Spurling's test
- ☐ Not literate in the English language
- ☐ Unable to complete the self-report functional questionnaires

☐ Subject meets BOTH inclusion criteria AND does not meet ANY of the exclusion criteria

☐ Subject does NOT meet inclusion criteria OR meets one of the exclusion criteria

www.manaraa.com

Appendix D: FOTO FS

Physical Therapy at St. Lukes - Physical Therapy at St. Luke's - West End

INTAKE FUNCTIONAL STATUS SUMMARY

Patient:
ID#
Date of Birth:
Initial DOS:
Body Part:
Impairment:
Surgery Type:
Surgery Date:

Risk-Adjustment Criteria	
Care Type:	**Gender:**
Body Part:	**Comorbidities:**
Severity:	**Payer:**
Age:	**Fear Avoidance:**
Acuity:	**Surgery Status:**

Functional Status Measures:	Intake Score	Interpretation of FS Scores/Stages Value
Patient's Physical FS Primary Measure		Patient's intake functional measure is out of (higher number = greater function).This FS measure places the patient in **Stage 4** and means the patient has good shoulder function.
Risk Adjusted Statistical FOTO*		Given the patient's risk-adjustment variables, like-patients nationally had a FS score of 52, **Stage 3**, at intake.

MCII = Points of change that is important to the patient)

MDC = Represents the smallest threshold to identify points of change that is greater than measurement error)

Rehabilitation Resource Predictor*	Predicted Value	Interpretation of Predicted Value
Points of Physical FS Change		
Discharge FS Score		Given this patient's risk-adjustment variables, and the actual intake FS score, FOTO predicts this patient will experience at least an increase in function of 6 points (to 73 or higher), putting them in the **Stage 4** level or higher at discharge.
Visits per Episode		
Duration of Episodes in Days		
Satisfaction Score		

Stage: 1 Exceedingly limited shoulder
Stage: 2 Poor shoulder
Stage: 3 Fair shoulder
Stage: 4 Good shoulder
Stage: 5 Excellent shoulder

* The above predictions are calculated for
1) patients who have previously utilized rehabilitation services from FOTO's national aggregate database and
2) using sophisticated analyses to risk adjust for the impact of ten important variables known to influence outcomes including Care type, Body Part/Impairment, Severity, Age, Acuity, Gender, Surgery, Fear Avoidance, Payer, and Comorbidities.

What Does This Mean For Improving Function

This chart displays the patient responses to the functional activities contained in the intake survey that generated the intake FS score. The activities are presented in the descending order of difficulty. Responses listed in the Intake column are the survey item levels of ability at intake. Given the change experienced by the comparative risk adjusted group in FOTO's data, it is anticipated the patient should be able to do the activities at the level indicated in the predicted column or higher at the completion of care, to place the patient in the predicted **Stage 4** functional level by discharge.

Patient responses to functional health questions that indicate dysfunction were as follows:

Activity (Question)	Amount of Limitation (Response) at Intake	Amount of Limitation (Response) predicted	Functional Limitation
How much difficulty do you have using your affected arm to place a 50 lb. box on a shelf overhead?			Other PT/OT Primary - G8990
How much difficulty do you have using your affected arm to place a 25 lb. box on a shelf overhead?			Other PT/OT Primary - G8990
Work overhead for more than 2 minutes?			Other PT/OT Primary - G8990

Produced and © 2001-2016 by
Focus On Therapeutic Outcomes, Inc.

Printed:

www.manaraa.com

INTAKE FUNCTIONAL STATUS SUMMARY

Patient: **Primary Body Part:** Shoulder **Initial DOS:**

Patient responses to functional health questions that indicate dysfunction were as follows:

Touch an object on the back seat while sitting in the front seat of a car?			Carrying, Moving & Handling Objects - G8984 Other PT/OT Primary - G8990
Reach an overhead shelf?			Other PT/OT Primary - G8990
Reach a shelf that is at shoulder height?			Carrying, Moving & Handling Objects - G8984

If the patient reaches the anticipated level on the above activities, other **Stage 4** activities the patient should be able to perform include:

- Adjusting the back of your collar with your affected hand - No difficulty
- Combing or brushing your hair using your affected arm - No difficulty
- Pull a medium weight object (5-10 lbs) from under a bed - No difficulty
- Move a heavy skillet (eg, cast iron skillet) from one stove burner to another - No difficulty
- Steady a jar while you loosen the jar lid - No difficulty
- Taking off glasses or sunglasses using your affected arm - No difficulty
- Place a can of soup (1 lb) on a shelf at shoulder height - No difficulty
- Reaching across to the middle of the table with your affected arm to get a salt shaker while sitting - No difficulty
- Turn a steering wheel in the opposite direction as your affected arm (eg, turn left if it is your right shoulder that is affected) - No difficulty
- Reach and pull the string that controls a light or fan - No difficulty

Additional Intake Information Gathered for the Clinician

- **Physician Referral:** Insurance Referral:
- **Patient reports other health problems as:**
- **BMI:**
- **Exercise prior to onset:**
- **Prescription medicine:**
- **Surgery:**
- **Fear avoidance belief about physical activity:**

Additional Surveys

	Intake	Scale
Physical Fear		

Physical Fear

Physical Fear Results:

	Fear Avd Belief About Phys Activ
Intake	

Crosswalk

	Intake	Scale
DASH		

Mathematical crosswalk from the Shoulder FS score to the DASH. For the DASH, a higher score indicates greater disability.

Produced and © 2001-2016 by
Focus On Therapeutic Outcomes, Inc.

Printed:

www.manaraa.com

Physical Therapy at St. Lukes - Physical Therapy at St. Luke's - West End

INTAKE FUNCTIONAL STATUS SUMMARY

Patient: **Primary Body Part:** Shoulder **Initial DOS:**

Pain Assessment Summary

Intensity

In the last 24 hours the level of pain was rated at: /10
In the last 30 days, the level of least pain was rated at: /10
and the level of most pain was rated at: /10

Character

Qualities of Pain
Patient reports that the pain feels

The intensity is

Influence of Activity
The pain is increased by

The pain is reduced by

CMS G-Codes

FOTO Shoulder Survey
CMS G-Code Options**

Functional Limitations Assessed in FOTO Shoulder Survey

Current Status	Goal Status	D/C** Status	Asked	Descriptor
G8984	G8985	G8986	2	Carrying, moving & handling objects functional limitation
G8987	G8988	G8989	0	Self care functional limitation
G8990	G8991	G8992	5	Other physical or occupational primary functional limitation

**Only report if this is a one time visit

CMS Impairment/Limitation/Restriction for FOTO Shoulder Survey

	Status	Limitation	G-Code	CMS Severity Modifier
Intake				
Predicted				

*Based on FOTO predicted change score

× _____

Clinician:

* Mean, Risk Adjusted, Intake Composite FS measures from FOTO aggregate database.
** As indicated by the ICF assignments to the survey items in the FOTO survey used.

Produced and © 2001-2016 by
Focus On Therapeutic Outcomes, Inc.

Printed:

www.manaraa.com

Appendix E: PSS/ASES

PENN/ASES SHOULDER SCORE

Subject #

Dominant Hand:	Gender:	Affected Arm:
L R Both (circle one)	M F (circle one)	L R Both (circle one)

PENN SHOULDER SCORE
Part I: Pain & Satisfaction: Please circle the number closest to your level of pain or satisfaction

	office use only
How bad is your pain today? 0 1 2 3 4 5 6 7 8 9 10 No Pain … Worst Pain Possible	(10 - #circled = __ x 5= __) ASES Pain
Pain at rest with your arm by your side: 0 1 2 3 4 5 6 7 8 9 10 No Pain … Worst Pain Possible	(10 - # circled)
Pain with normal activities (eating, dressing, bathing): 0 1 2 3 4 5 6 7 8 9 10 No Pain … Worst Pain Possible	(10 - # circled) (score "0" if not applicable)
Pain with strenuous activities (reaching, lifting, pushing, pulling, throwing): 0 1 2 3 4 5 6 7 8 9 10 No Pain … Worst Pain Possible	(10 - # circled) (score "0" if not applicable)
PAIN SCORE:	=___/30
How satisfied are you with the current level of function of your shoulder? 0 1 2 3 4 5 6 7 8 9 10 Not Satisfied … Very Satisfied	=___/10 (# circled)

PLEASE TURN OVER TO COMPLETE QUESTIONNAIRE

OFFICE USE ONLY

Today's Date: / /

	PENN SHOULDER SCORE (PSS)	ASES SHOULDER SCORE (ASES)
Pain	/30	/50
Satisfaction	/10	
Function	/60	/50
TOTAL	/100	/100

© 1999 Brian G. Leggin
**The author grants unrestricted use of this questionnaire for patient care and clinical research purposes.

www.manaraa.com

PENN SHOULDER SCORE/ASES SHOULDER SCORE Part II: Function: Please circle the number that best describes the level of difficulty you might have performing each activity.	No difficulty	Some difficulty	Much difficulty	Can't do at all	Did not do before injury
1. Reach the small of your back to tuck in your shirt with your hand.	3	2	1	0	X
2. Wash the middle of your back/hook bra. (ASES #3)	3	2	1	0	X
3. Perform necessary toileting activities. (ASES #4)	3	2	1	0	X
4. Wash the back of opposite shoulder.	3	2	1	0	X
5. Comb hair. (ASES # 5)	3	2	1	0	X
6. Place hand behind head with elbow held straight out to the side.	3	2	1	0	X
7. Dress self (including put on coat and pull shirt of overhead. (ASES # 1)	3	2	1	0	X
8. Sleep on affected side. (ASES # 2)	3	2	1	0	X
9. Open a door with affected side.	3	2	1	0	X
10. Carry a bag of groceries with affected arm.	3	2	1	0	X
11. Carry a briefcase/small suitcase with affected arm.	3	2	1	0	X
12. Place a soup can (1-2 lbs.) on a shelf at shoulder level without bending elbow.	3	2	1	0	X
13. Place a one gallon container (8-10 lbs.) on a shelf at shoulder level without bending elbow.	3	2	1	0	X
14. Reach a shelf above your head without bending your elbow. (ASES # 6)	3	2	1	0	X
15. Place a soup can (1-2 lbs.) on a shelf overhead without bending your elbow.	3	2	1	0	X
16. Place a one gallon container (8-10 lbs.) on a shelf Overhead without bending your elbow. (ASES # 7)	3	2	1	0	X
17. Perform usual sport/hobby. (ASES # 9)	3	2	1	0	X
18. Perform household chores (cleaning, laundry, cooking).	3	2	1	0	X
19. Throw overhand/swim/overhead racquet sports. (circle all that apply to you) (ASES # 8)	3	2	1	0	X
20. Work full-time at your regular job. (ASES # 10)	3	2	1	0	X
SCORING: (office use only)	PSS	PSS	PSS	PSS	PSS
PSS Total of all columns = _____ (a)	_____	_____	_____	_____	_____
Number of "X's" x 3 =_____ (b), 60 – _____ (b) = _____ (c)					
(If no "X's" are circled, function score = total of columns)	ASES	ASES	ASES	ASES	
PSS Function Score = ____(a) ÷____(c) = ____ x 60 = ____ of 60	_____	_____	_____	_____	
ASES Score = _____					
Total of shaded columns: ____ x 5/3 = _____ of 60					

© 1999 Brian G. Leggin **The author grants unrestricted use of this questionnaire for patient care and clinical research purposes.

www.manaraa.com

Appendix F: Examination

For Office Use Only:
Subject #_____

Pain At Night or At Rest

- ☐ Constantly
- ☐ Intermittently
- ☐ None

AROM compared to PROM

- ☐ AROM less than (<) PROM
- ☐ AROM nearly equal to (≈) PROM
- ☐ AROM equal to (=) PROM

Pain reproduction with ROM

- ☐ Prior to end range
- ☐ At end range
- ☐ None or with overpressure at end range

www.manaraa.com

Appendix G: Shoulder Symptom Irritability Classification

For Office Use Only:
Subject #_____

PHYSICAL THERAPY AT StLuke's

Irritability Rating

Based upon your examination of this patient, please rate the level of shoulder symptom irritability:

☐ High ☐ Moderate ☐ Low

☐ Unclassifiable (Reason:_____)

www.manaraa.com

Appendix H: Treatment Strategy

For Office Use Only:
Subject #_____

Which of the below treatments strategies would

BEST be used for this patient TODAY?

(Select only one)

☐ Provide moderate–high physical stress
Address specific impairments
Restore of high-demand functional activity

☐ Provide mild–moderate physical stress
Address specific impairments
Restore basic-level functional activity

☐ Minimize Physical Stress
Modify activities
Monitor impairments

www.manaraa.com

Appendix I: Intervention Intensity

For Office Use Only:
Subject #_____

Do you plan to provide any of the below treatments to this patient TODAY?

1. **Exercises addressing muscular weakness** ☐ Yes ☐ No
   If yes, select one option below:
   - ☐ Active range of motion (no external load)
   - ☐ Mild-moderate resistive strength training
   - ☐ Moderate-high resistive strength training

2. **Exercises addressing mobility** ☐ Yes ☐ No
   If yes, select one option below:
   - ☐ Range of motion exercises (non-end range stress; pain free)
   - ☐ Range of motion exercises (end range stress; transient or shorter hold times)
   - ☐ Range of motion exercises (end range stress; longer duration hold times)

3. **Shoulder joint mobilizations** ☐ Yes ☐ No
   If yes, select one option below:
   - ☐ Low grade; not achieving end range
   - ☐ High grade; achieving end range

4. **Electrical Agents for pain control (e.g. TENS)** ☐ Yes ☐ No

5. **Thermal modalities** ☐ Yes ☐ No
   If yes, select one option below:
   - ☐ For pain control or relaxation
   - ☐ To facilitate tissue extensibility

6. **Recommendations for Functional Activity** ☐ Yes ☐ No
   If yes, select one option below:
   - ☐ Avoid provocative functional activities
   - ☐ Encourage basic-level functional activities
   - ☐ Encourage high-demand functional activities

www.manaraa.com

Appendix J: Inter-rater Contingency Tables for Each Site

Site 1: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	0	1	0
	Moderate	0	5	0
	High	1	1	2

PABAK-OS = 0.55 (95% CI 0.25, 0.85)
Rater Agreement = 70%

Site 2: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	3	1	0
	Moderate	2	2	1
	High	0	3	2

PABAK-OS = 0.55 (95% CI 0.29, 0.80)
Rater Agreement = 50%

Site 3: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	4	1	0
	Moderate	1	4	0
	High	0	1	1

PABAK-OS = 0.78 (95% CI 0.50, 1.0)
Rater Agreement = 75%

Site 4: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	1	0	0
	Moderate	0	2	1
	High	0	2	2

PABAK-OS = 0.66 (95% CI 0.33, 1.0)
Rater Agreement = 63%

www.manaraa.com

Site 5: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	0	1	0
	Moderate	0	0	1
	High	0	0	0

PABAK-OS = 0.09 (95% CI 0, 0.77)
Rater Agreement = 0%

Site 6: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	9	3	0
	Moderate	1	9	1
	High	0	1	4

PABAK-OS = 0.81 (95% CI 0.63, 0.99)
Rater Agreement = 79%

Site 7: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	1	1	0
	Moderate	0	4	0
	High	0	2	2

PABAK-OS = 0.73 (95% CI 0.43, 1.0)
Rater Agreement = 70%

Site 8: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	3	0	0
	Moderate	0	1	0
	High	0	0	2

PABAK-OS = 1.0 (95% CI 0.61, 1.0)
Rater Agreement = 100%

www.manaraa.com

Site 9: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	0	0	0
	Moderate	0	2	1
	High	0	0	0

PABAK-OS = 0.70 (95% CI 0.15, 1.0)
Rater Agreement = 67%

Site 10: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	2	0	0
	Moderate	1	0	1
	High	0	0	0

PABAK-OS = 0.55 (95% CI 0.08, 1.0)
Rater Agreement = 50%

Site 11: Inter-Rater Reliability of Shoulder Symptom Irritability Classification

		Rater 2		
		Low	**Moderate**	**High**
Rater 1	**Low**	0	1	0
	Moderate	0	1	0
	High	0	2	0

PABAK-OS = 0.32 (95% CI 0, 0.80)
Rater Agreement = 25%

www.manaraa.com

Appendix K: Contingency Tables for Shoulder Symptom Irritability and Treatment Strategy

Rater 1: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	1	0
	Moderate	0	2	1
	High	0	0	2

PABAK-OS = 0.70 (95% CI 0.31, 1.00)
Agreement = 67%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 9: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	4	0	0
	Moderate	0	6	0
	High	0	0	2

PABAK-OS = 1.00 (95% CI 0.73, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 3: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	2	0
	Moderate	0	5	1
	High	0	0	2

PABAK-OS = 0.73 (95% CI 0.43, 1.00)
Agreement = 80%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Rater 4: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	1	2	0
	Moderate	0	1	0
	High	0	1	1

PABAK-OS = 0.55 (95% CI 0.16, 0.94)
Agreement = 50%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 5: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	4	0	0
	Moderate	0	5	0
	High	0	0	5

PABAK-OS = 1.00 (95% CI 0.75, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 6: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	0	0
	Moderate	0	1	0
	High	0	1	0

PABAK-OS = 0.55 (95% CI 0, 1.00)
Agreement = 50%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Rater 7: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	1	0
	Moderate	0	2	1
	High	0	0	3

PABAK-OS = 0.74 (95% CI 0.39, 1.00)
Agreement = 71%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 8: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	1	0
	Moderate	0	5	0
	High	0	0	1

PABAK-OS = 0.87 (95% CI 0.51, 1.00)
Agreement = 86%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 9: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	0	0
	Moderate	0	2	0
	High	0	0	1

PABAK-OS = 1.00 (95% CI 0.45, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Rater 10: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	2	0	0
	Moderate	0	1	0
	High	0	0	1

PABAK-OS = 1.00 (95% CI 0.53, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 11: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	7	2	0
	Moderate	0	12	0
	High	0	2	3

PABAK-OS = 0.86 (95% CI 0.68, 1.00)
Agreement = 85%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 12: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	3	1	1
	Moderate	0	5	1
	High	0	0	3

PABAK-OS = 0.68 (95% CI 0.53, 0.93)
Agreement = 79%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Rater 13: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	1	0
	Moderate	0	0	1
	High	0	0	0

PABAK-OS = 0.10 (95% CI 0, 0.77)
Agreement = 0%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 14: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	4	6	0
	Moderate	0	9	0
	High	0	0	5

PABAK-OS = 0.78 (95% CI 0.58, 0.97)
Agreement = 75%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 15: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	0	0
	Moderate	0	4	0
	High	0	0	0

PABAK-OS = 1.00 (95% CI 0.53, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Rater 16: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	1	2	0
	Moderate	0	1	0
	High	0	0	0

PABAK-OS = 0.55 (95% CI 0.08, 1.00)
Agreement = 50%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 17: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	2	4	0
	Moderate	0	5	0
	High	0	1	0

PABAK-OS = 0.62 (95% CI 0.35, 0.90)
Agreement = 58%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 18: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	0	0
	Moderate	0	3	0
	High	0	0	0

PABAK-OS = 1.00 (95% CI 0.45, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Rater 19: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	2	1	0
	Moderate	0	1	0
	High	0	0	2

PABAK-OS = 0.85 (95% CI 0.46, 1.00)
Agreement = 83%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 20: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	3	0	0
	Moderate	0	3	0
	High	0	0	0

PABAK-OS = 1.00 (95% CI 0.61, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 21: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	0	0
	Moderate	0	6	0
	High	0	1	3

PABAK-OS = 0.91 (95% CI 0.61, 1.00)
Agreement = 90%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

Rater 22: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	0	1	0
	Moderate	0	1	0
	High	0	1	1

PABAK-OS = 0.55 (95% CI 0.08, 1.0)
Agreement = 50%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 23: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	1	0	0
	Moderate	0	3	0
	High	0	0	4

PABAK-OS = 1.00 (95% CI 0.67, 1.00)
Agreement = 100%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

Rater 24: Agreement between selected Shoulder Symptom Irritability and Treatment Strategy

		Intensity		
		High	**Moderate**	**Low**
SSIC	**Low**	1	0	0
	Moderate	0	3	1
	High	0	0	3

PABAK-OS = 0.90 (95% CI 0.58, 1.00)
Agreement = 89%
SSIC, Shoulder Symptom Irritability Classification; Intensity, Intensity of Treatment Strategy

www.manaraa.com

REFERENCES

1. Luime JJ, Koes BW, Hendriksen IJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol.* 2004;33(2):73-81.
2. McClure PW, Michener LA. Staged Approach for Rehabilitation Classification: Shoulder Disorders (STAR-Shoulder). *Phys Ther.* 2015;95(5):791-800.
3. Van der Heijden GJ. Shoulder disorders: a state-of-the-art review. *Baillieres Best Pract Res Clin Rheumatol.* 1999;13:287-309.
4. Macfarlane GJ, Hunt IM, Silman AJ. Predictors of chronic shoulder pain: a population based prospective study. *J Rheumatol.* 1998;25(8):1612-1615.
5. Ludewig PM, Lawrence RL, Braman JP. What's in a Name? Using movement system diagnoses versus pathoanatomic diagnoses. *JOrthopSports Phys Ther.* 2013;43(5):280-283.
6. Miller-Spoto M, Gombatto SP. Diagnostic labels assigned to patients with orthopedic conditions and the influence of the label on selection of interventions: a qualitative study of orthopaedic clinical specialists. *Phys Ther.* 2014;94(6):776-791.
7. Brody LT. Effective therapeutic exercise prescription: the right exercise at the right dose. *J Hand Ther.* 2012;25(2):220-231; quiz 232.
8. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther.* 1995;75(6):470-485; discussion 485-479.
9. Maitland GD, Banks K, English K, Hengveld E, eds. *Maitland's Vertebral Manipulation.* 6th ed. London: Butterworth; 2001.
10. Mueller MJ, Maluf KS. Tissue adaptation to physical stress: a proposed "Physical Stress Theory" to guide physical therapist practice, education, and research. *Phys Ther.* 2002;82(4):383-403.
11. Maitland GD, Edwards BC. *Vertebral manipulation.* 5th ed. London ; Boston: Butterworths; 1986.
12. Maitland GD. *Peripheral manipulation.* 3rd ed. London ; Boston: Butterworth-Heinemann; 1991.
13. Barakatt ET. *Low back pain irritability: Current application and development of a theoretical construct* [Ph.D.]. Ann Arbor, University of California, Davis; 2004.
14. Kelley MJ, McClure PW, Leggin BG. Frozen shoulder: evidence and a proposed model guiding rehabilitation. *J Orthop Sports Phys Ther.* 2009;39(2):135-148.
15. Barakatt ET, Romano PS, Riddle DL, Beckett LA, Kravitz R. An Exploration of Maitland's Concept of Pain Irritability in Patients with Low Back Pain. *J Man Manip Ther.* 2009;17(4):196-205.
16. Barakatt ET, Romano PS, Riddle DL, Beckett LA. The Reliability of Maitland's Irritability Judgments in Patients with Low Back Pain. *J Man Manip Ther.* 2009;17(3):135-140.
17. Delitto A, George SZ, Van Dillen LR, et al. Low back pain. *J Orthop Sports Phys Ther.* 2012;42(4):A1-57.
18. Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute

www.manaraa.com

low back pain: a randomized clinical trial. *Spine (Phila Pa 1976).* 2003;28(13):1363-1371; discussion 1372.
19. Wainner RS, Fritz JM, Irrgang JJ, Boninger ML, Delitto A, Allison S. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine (Phila Pa 1976).* 2003;28(1):52-62.
20. Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. *Spine (Phila Pa 1976).* 2006;31(6):623-631.
21. Fritz JM, Cleland JA, Childs JD. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *J Orthop Sports Phys Ther.* 2007;37(6):290-302.
22. Apeldoorn AT, Ostelo RW, van Helvoirt H, et al. A randomized controlled trial on the effectiveness of a classification-based system for subacute and chronic low back pain. *Spine (Phila Pa 1976).* 2012;37(16):1347-1356.
23. Apeldoorn AT, Bosmans JE, Ostelo RW, de Vet HC, van Tulder MW. Cost-effectiveness of a classification-based system for sub-acute and chronic low back pain. *Eur Spine J.* 2012;21(7):1290-1300.
24. Stanton TR, Fritz JM, Hancock MJ, et al. Evaluation of a treatment-based classification algorithm for low back pain: a cross-sectional study. *Phys Ther.* 2011;91(4):496-509.
25. Stanton TR, Hancock MJ, Apeldoorn AT, Wand BM, Fritz JM. What Characterizes People Who Have an Unclear Classification Using a Treatment-Based Classification Algorithm for Low Back Pain? A Cross-Sectional Study. *Phys Ther.* 2012.
26. Stanton TR, Hancock MJ, Apeldoorn AT, Wand BM, Fritz JM. What characterizes people who have an unclear classification using a treatment-based classification algorithm for low back pain? A cross-sectional study. *Phys Ther.* 2013;93(3):345-355.
27. Kelley MJ, Shaffer MA, Kuhn JE, et al. Shoulder pain and mobility deficits: adhesive capsulitis: clinical practice guidelines linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. *Journal of Orthopaedic & Sports Physical Therapy.* 2013;43(5):A1-A31.
28. Maitland GD. *Vertebral manipulation.* 2nd ed. London,: Butterworths; 1968.
29. Childs JD, Fritz JM, Wu SS, et al. Implications of early and guideline adherent physical therapy for low back pain on utilization and costs. *BMC health services research.* 2015;15(1):150.
30. Fritz JM, Cleland JA, Brennan GP. Does adherence to the guideline recommendation for active treatments improve the quality of care for patients with acute low back pain delivered by physical therapists? *Med Care.* 2007;45(10):973-980.
31. Dempsey AL, Mills T, Karsch RM, Branch TP. Maximizing total end range time is safe and effective for the conservative treatment of frozen shoulder patients. *Am J Phys Med Rehabil.* 2011;90(9):738-745.
32. Diercks RL, Stevens M. Gentle thawing of the frozen shoulder: a prospective study of supervised neglect versus intensive physical therapy in seventy-seven

www.manaraa.com

patients with frozen shoulder syndrome followed up for two years. *Journal of Shoulder and Elbow Surgery.* 2004;13(5):499-502.
33. IACAT. International Association of Computerized Adaptive Testing. 2016; http://www.iacat.org/. Accessed 27 January 2016.
34. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum.* 2009;61(5):623-632.
35. Badalamente M, Coffelt L, Elfar J, et al. Measurement scales in clinical research of the upper extremity, part 2: outcome measures in studies of the hand/wrist and shoulder/elbow. *The Journal of hand surgery.* 2013;38(2):407-412.
36. APTA. *Physical Therapists & The Performance of Dry Needling.* 2012.
37. Green S, Buchbinder R, Hetrick S. Physiotherapy interventions for shoulder pain (Review). 2013; http://summaries.cochrane.org/CD004258/some-physiotherapy-interventions-are-effective-for-shoulder-pain-in-some-cases.
38. Reinold MM, Macrina LC, Wilk KE, Dugas JR, Cain EL, Andrews JR. The Effect of Neuromuscular Electrical Stimulation of the Infraspinatus on Shoulder External Rotation Force Production After Rotator Cuff Repair Surgery. *The American Journal of Sports Medicine.* 2008;36(12):2317-2321.
39. Baker LL, Parker K. Neuromuscular Electrical Stimulation of the Muscles Surrounding the Shoulder. *Physical Therapy.* 1986;66(12):1930-1937.
40. Hayes K, Petersen C. Reliability of assessing end-feel and pain and resistance sequence in subjects with painful shoulders and knees. *The Journal of orthopaedic and sports physical therapy.* 2001;31(8):432.
41. Chesworth BM, MacDermid JC, Roth JH, Patterson SD. Movement diagram and "end-feel" reliability when measuring passive lateral rotation of the shoulder in patients with shoulder pathology. *Physical therapy.* 1998;78(6):593-601.
42. McClure PW, Greenberg E, Kareha SM. Evaluation and management of scapular dysfunction. *Sports Med Arthrosc.* 2012;20(1):39-48.
43. Tate AR, McClure PW, Young IA, Salvatori R, Michener LA. Comprehensive impairment-based exercise and manual therapy intervention for patients with subacromial impingement syndrome: a case series. *J Orthop Sports Phys Ther.* 2010;40(8):474-493.
44. Escamilla RF, Hooks TR, Wilk KE. Optimal management of shoulder impingement syndrome. *Open Access Journal of Sports Medicine.* 2014;5:13-24.
45. Hart DL, Werneke MW, George SZ, et al. Screening for elevated levels of fear-avoidance beliefs regarding work or physical activities in people receiving outpatient therapy. *Phys Ther.* 2009;89(8):770-785.
46. Kuhn JE. Exercise in the treatment of rotator cuff impingement: a systematic review and a synthesized evidence-based rehabilitation protocol. *Journal of shoulder and elbow surgery.* 2009;18(1):138-160.
47. Schellingerhout JM, Verhagen AP, Thomas S, Koes BW. Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. *Man Ther.* 2008;13(6):478-483.
48. Somerville LE, Willits K, Johnson AM, et al. Clinical Assessment of Physical Examination Maneuvers for Rotator Cuff Lesions. *Am J Sports Med.* 2014;42(8):1911-1919.

www.manaraa.com

49. Ostor AJ, Richards CA, Tytherleigh-Strong G, et al. Validation of clinical examination versus magnetic resonance imaging and arthroscopy for the detection of rotator cuff lesions. *Clin Rheumatol.* 2013;32(9):1283-1291.
50. Hegedus EJ, Goode AP, Cook CE, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med.* 2012;46(14):964-978.
51. Collin P, Treseder T, Denard PJ, Neyton L, Walch G, Ladermann A. What is the Best Clinical Test for Assessment of the Teres Minor in Massive Rotator Cuff Tears? *Clin Orthop Relat Res.* 2015;473(9):2959-2966.
52. Bruyn GA, Pineda C, Hernandez-Diaz C, et al. Validity of ultrasonography and measures of adult shoulder function and reliability of ultrasonography in detecting shoulder synovitis in patients with rheumatoid arthritis using magnetic resonance imaging as a gold standard. *Arthritis Care Res (Hoboken).* 2010;62(8):1079-1086.
53. Razmjou H, Fournier-Gosselin S, Christakis M, Pennings A, ElMaraghy A, Holtby R. Accuracy of magnetic resonance imaging in detecting biceps pathology in patients with rotator cuff disorders: comparison with arthroscopy. *J Shoulder Elbow Surg.* 2016;25(1):38-44.
54. Phillips JC, Cook C, Beaty S, Kissenberth MJ, Siffri P, Hawkins RJ. Validity of noncontrast magnetic resonance imaging in diagnosing superior labrum anterior-posterior tears. *J Shoulder Elbow Surg.* 2013;22(1):3-8.
55. Mohtadi NG, Vellet AD, Clark ML, et al. A prospective, double-blind comparison of magnetic resonance imaging and arthroscopy in the evaluation of patients presenting with shoulder pain. *J Shoulder Elbow Surg.* 2004;13(3):258-265.
56. Seo JB, Yoo JS, Jang HS, Kim JS. Correlation of clinical symptoms and function with fatty degeneration of infraspinatus in rotator cuff tear. *Knee Surg Sports Traumatol Arthrosc.* 2015;23(5):1481-1488.
57. Pennington RG, Bottomley NJ, Neen D, Brownlow HC. Radiological features of osteoarthritis of the acromiclavicular joint and its association with clinical symptoms. *Journal of orthopaedic surgery (Hong Kong).* 2008;16(3):300-302.
58. Kuhn JE, Dunn WR, Sanders R, et al. Effectiveness of physical therapy in treating atraumatic full-thickness rotator cuff tears: a multicenter prospective cohort study. *J Shoulder Elbow Surg.* 2013;22(10):1371-1379.
59. Lahteenmaki HE, Virolainen P, Hiltunen A, Heikkila J, Nelimarkka OI. Results of early operative treatment of rotator cuff tears with acute symptoms. *J Shoulder Elbow Surg.* 2006;15(2):148-153.
60. Dunn WR, Kuhn JE, Baumgarten K, et al. Defining Indications for Rotator Cuff Repair: Predictors of Failure of Nonoperative Treatment of Chronic, Symptomatic, Full-Thickness Rotator Cuff Tears. *Journal of Shoulder and Elbow Surgery.* 2013;22(4):e28.
61. Delitto A, George SZ, Van Dillen L, et al. Low Back Pain Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. *JOURNAL OF ORTHOPAEDIC & SPORTS PHYSICAL THERAPY.* 2012;42(4):A1-A57.

www.manaraa.com

62. Childs JD, Fritz JM, Piva SR, Whitman JM. Proposal of a classification system for patients with neck pain. *J Orthop Sports Phys Ther.* 2004;34(11):686-696; discussion 697-700.
63. Childs JD, Cleland JA, Elliott JM, et al. Neck pain: Clinical practice guidelines linked to the International Classification of Functioning, Disability, and Health from the Orthopedic Section of the American Physical Therapy Association. *J Orthop Sports Phys Ther.* 2008;38(9):A1-A34.
64. Alrwaily M, Timko M, Schneider M, et al. Treatment-Based Classification System for Low Back Pain: Revision and Update. *Phys Ther.* 2016;96(7):1057-1066.
65. Wang WT, Olson SL, Campbell AH, Hanten WP, Gleeson PB. Effectiveness of physical therapy for patients with neck pain: an individualized approach using a clinical decision-making algorithm. *American journal of physical medicine & rehabilitation.* 2003;82(3):203-218.
66. Fritz JM, Brennan GP. Preliminary examination of a proposed treatment-based classification system for patients receiving physical therapy interventions for neck pain. *Phys Ther.* 2007;87(5):513-524.
67. Tashjian RZ. Epidemiology, natural history, and indications for treatment of rotator cuff tears. *Clin Sports Med.* 2012;31(4):589-604.
68. Matsen FA, 3rd, Chebli C, Lippitt S. Principles for the evaluation and management of shoulder instability. *J Bone Joint Surg Am.* 2006;88(3):648-659.
69. Denard PJ, Wirth MA, Orfaly RM. Management of glenohumeral arthritis in the young adult. *J Bone Joint Surg Am.* 2011;93(9):885-892.
70. Monk AP, Garfjeld Roberts P, Logishetty K, et al. Evidence in managing traumatic anterior shoulder instability: a scoping review. *Br J Sports Med.* 2013.
71. Geoffrey D. Maitland, 1924–2010. *Physical Therapy.* 2010;90(3):326-326.
72. Hengeveld E, Banks K, Maitland GD. *Maitland's peripheral manipulation.* 4th ed. Edinburgh ; New York: Elsevier/Butterworth Heinemann; 2005.
73. Smart K, Doody C. The clinical reasoning of pain by experienced musculoskeletal physiotherapists. *Man Ther.* 2007;12(1):40-49.
74. Lentz TA, Beneciuk JM, Bialosky JE, et al. Development of a Yellow Flag Assessment Tool for Orthopaedic Physical Therapists: Results From the Optimal Screening for Prediction of Referral and Outcome (OSPRO) Cohort. *J Orthop Sports Phys Ther.* 2016;46(5):327-343.
75. Kareha SM, Buggey JT, Miller TW, Carolan GF. Reliability of Shoulder Irritability Levels: A Pilot Study. *J Orthop Sports Phys Ther.* 2015;45(1):A103.
76. McFarland EG, Garzon-Muvdi J, Jia X, Desai P, Petersen SA. Clinical and diagnostic tests for shoulder disorders: a critical review. *British journal of sports medicine.* 2010;44(5):328-332.
77. APTA. Interactive Guide to Physical Therapist Practice. In:2003.
78. Sahrmann SA. Diagnosis by the physical therapist--a prerequisite for treatment. A special communication. *Phys Ther.* 1988;68(11):1703-1706.
79. May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. *Physiotherapy.* 2010;96(3):179-190.

www.manaraa.com

80. Meister K, Andrews JR. Classification and treatment of rotator cuff injuries in the overhand athlete. *J Orthop Sports Phys Ther.* 1993;18(2):413-421.
81. Bot S, Terwee C, Van der Windt D, Bouter L, Dekker J, De Vet H. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Annals of the Rheumatic Diseases.* 2004;63(4):335-341.
82. Hart DL, Wang YC, Cook KF, Mioduski JE. A computerized adaptive test for patients with shoulder impairments produced responsive measures of function. *Phys Ther.* 2010;90(6):928-938.
83. Parker R, Vannest K, Davis J. Inter-Rater Reliability for Multi-Category Scales: Problems and a Solution in PABAK-OS; Unpublished Work. In:2011.
84. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy.* 2005;85(3):257-268.
85. Hart DL. The power of outcomes: FOTO Industrial Outcomes Tool -- Initial assessment. *Work.* 2001;16(1):39-51.
86. Dobrzykowski EA, Nance T. The Focus On Therapeutic Outcomes (FOTO) outpatient orthopedic rehabilitation database: results of 1994-1996. *Journal of Rehabilitation Outcomes Measurement.* 1997;1:56-60.
87. Atkinson HL, Nixon-Cave K. A tool for clinical reasoning and reflection using the international classification of functioning, disability and health (ICF) framework and patient management model. *Phys Ther.* 2011;91(3):416-430.
88. Crane PK, Hart DL, Gibbons LE, Cook KF. A 37-item shoulder functional status item pool had negligible differential item functioning. *J Clin Epidemiol.* 2006;59(5):478-484.
89. Hart DL, Cook KF, Mioduski JE, Teal CR, Crane PK. Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *J Clin Epidemiol.* 2006;59(3):290-298.
90. Cook KF, Roddey TS, Gartsman GM, Olson SL. Development and psychometric evaluation of the Flexilevel Scale of Shoulder Function. *Medical Care.* 2003;41(7):823-835.
91. Cook KF, Roddey TS, O'Malley KJ, Gartsman GM. Development of a Flexilevel Scale for use with computer-adaptive testing for assessing shoulder function. *J Shoulder Elbow Surg.* 2005;14(1 Suppl S):90S-94S.
92. Wang YC, Hart DL, Cook KF, Mioduski JE. Translating shoulder computerized adaptive testing generated outcome measures into clinical practice. *J Hand Ther.* 2010;23(4):372-382; quiz 383.
93. Leggin BG, Lannotti J. Shoulder outcome measurement. In: Lannotti JP, Williams GR, eds. *Disorders of the Shoulder: Diagnosis and Management.* Philadelphia, PA: Lippincott, Williams & Wilkins; 1999:1024-1040.
94. Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP, Williams GR, Jr. The Penn shoulder score: reliability and validity. *J Orthop Sports Phys Ther.* 2006;36(3):138-151.
95. Richards RR, An K-N, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *Journal of Shoulder and Elbow Surgery.* 1994;3(6):347-352.

www.manaraa.com

96. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. *J Shoulder Elbow Surg.* 2002;11(6):587-594.
97. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis care & research.* 2011;63(S11):S174-S188.
98. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. *The Journal of Bone & Joint Surgery.* 2010;92(2):296-303.
99. St-Pierre C, Desmeules F, Dionne CE, Fremont P, MacDermid JC, Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disabil Rehabil.* 2016;38(2):103-122.
100. Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. *J Shoulder Elbow Surg.* 2009;18(6):920-926.
101. Smidt N, van der Windt DA, Assendelft WJ, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil.* 2002;83(8):1145-1150.
102. Berthier F, Potel G, Leconte P, Touze MD, Baron D. Comparative study of methods of measuring acute pain intensity in an ED. *Am J Emerg Med.* 1998;16(2):132-136.
103. Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. *Aust J Physiother.* 2001;47(4):289-294.
104. Riddle DL, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting. Shoulder measurements. *Phys Ther.* 1987;67(5):668-673.
105. Williams JG, Callagan M. Comparison of visual estimation of goniometry in determination of a shoulder joint angle. *Physiotherapy.* 1990;76:655-657.
106. Hoving JL, Buchbinder R, Green S, et al. How reliably do rheumatologists measure shoulder movement? *Ann Rheum Dis.* 2002;61(7):612-616.
107. Kareha SM. Shoulder Symptom Irritability Scale Rater Training. In:2016.
108. Fuentes J, Armijo-Olivo S, Funabashi M, et al. Enhanced therapeutic alliance modulates pain intensity and muscle pain sensitivity in patients with chronic low back pain: an experimental controlled study. *Physical therapy.* 2014;94(4):477-489.
109. Vannest KJ, Parker RI, Gonen O, Adiguzel T. Single Case Research: web based calculators for SCR analysis. (Version 2.0) [Web-based application]. 2016; www.singlecaseresearch.org. Accessed Monday 13th November 2017.

www.manaraa.com

110. Carlson M, Protsman L, Tomaka J. Graphic Organizers Can Facilitate Selection of Statistical Tests Part 1: Analysis of Group Differences. *Journal of Physical Therapy Education.* 2005;19(2).
111. Jensen GM, Shepard KF, Hack LM. The novice versus the experienced clinician: insights into the work of the physical therapist. *Physical Therapy.* 1990;70(5):314-323.
112. Jensen GM, Shepard KF, Gwyer J, Hack LM. Attribute dimensions that distinguish master and novice physical therapy clinicians in orthopedic settings. *Physical Therapy.* 1992;72(10):711-722.
113. Shepard KF, Hack LM, Gwyer J, Jensen GM. Describing expert practice in physical therapy. *Qualitative Health Research.* 1999;9(6):746-758.
114. APTA. Information for New Professionals. 2017; http://www.apta.org/NewProfessionals/. Accessed December 27, 2017.
115. APTA. *Physical Therapist Member Demographic Profile.* 2014.
116. ABPTS. ABPTS Certified Specialists Statistics. 2017; http://www.abpts.org/About/Statistics/. Accessed 2017 November 20.
117. Hill CL, Gill TK, Shanahan EM, Taylor AW. Prevalence and correlates of shoulder pain and stiffness in a population-based study: the North West Adelaide Health Study. *International journal of rheumatic diseases.* 2010;13(3):215-222.
118. Hardyck C, Petrinovich LF. Left-handedness. *Psychol Bull.* 1977;84(3):385-404.
119. van der Windt DA, Koes BW, de Jong BA, Bouter LM. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis.* 1995;54(12):959-964.
120. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
121. Fritz JM, Brennan GP, Clifford SN, Hunter SJ, Thackeray A. An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. *Spine.* 2006;31(1):77-82.
122. Razmjou H, Kramer JF, Yamada R. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low back pain. *Journal of Orthopaedic & Sports Physical Therapy.* 2000;30(7):368-389.
123. Kilpikoski S, Airaksinen O, Kankaanpää M, Leminen P, Videman T, Alen M. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine.* 2002;27(8):E207-E214.
124. Clare HA, Adams R, Maher CG. Reliability of McKenzie classification of patients with cervical or lumbar pain. *Journal of manipulative and physiological therapeutics.* 2005;28(2):122-127.
125. Riddle DL, Rothstein JM. Intertester reliability of McKenzie's classifications of the syndrome types present in patients with low back pain. *Spine.* 1993;18(10):1333-1344.
126. Werneke MW, Deutscher D, Hart DL, et al. McKenzie lumbar classification: inter-rater agreement by physical therapists with different levels of formal McKenzie postgraduate training. *Spine (Phila Pa 1976).* 2014;39(3):E182-190.
127. McClure P, Tate AR, Kareha SM, Irwin D, Zlupko E. A clinical method for identifying scapular dyskinesis, part 1: reliability. *J Athl Train.* 2009;44(2):160-164.

www.manaraa.com

128. Portney LG, Watkins MP. *Foundations of clinical research : applications to practice.* 3rd ed. Upper Saddle River, N.J.: Pearson/Prentice Hall; 2009.
129. Moxley JH, Ericsson KA, Scheiner A, Tuffiash M. The effects of experience and disuse on crossword solving. *Applied Cognitive Psychology.* 2015;29(1):73-80.
130. Hart DL, Dobrzykowski EA. Influence of orthopaedic clinical specialist certification on clinical outcomes. *J Orthop Sports Phys Ther.* 2000;30(4):183-193.
131. Resnik L, Hart DL. Using clinical outcomes to identify expert physical therapists. *Phys Ther.* 2003;83(11):990-1002.
132. Resnik L, Dobrzykowski E. Guide to outcomes measurement for patients with low back pain syndromes. *J Orthop Sports Phys Ther.* 2003;33(6):307-316; discussion 317-308.
133. Bradberry T, Greaves J. *Emotional Intelligence 2.0.* Enhanced, New edition (June 16, 2009) ed: TalentSmart; 2010.
134. Weng H-C. Does the physician's emotional intelligence matter?: Impacts of the physician's emotional intelligence on the trust, patient-physician relationship, and satisfaction. *Health Care Management Review.* 2008;33(4):280-288.
135. Adams KL, Iseler JI. The Relationship of Bedside Nurses' Emotional Intelligence With Quality of Care. *Journal of Nursing Care Quality.* 2014;29(2):174-181.
136. Wagner PJ, Moseley GC, Grant MM, Gore JR, Owens C. Physicians' emotional intelligence and patient satisfaction. *Family medicine.* 2002;34(10):750-754.
137. AAOS. Patient Reported Outcome Measures. 2016; https://www.aaos.org/Quality/Performance_Measures/Patient_Reported_Outcome_Measures/. Accessed December 23, 2017.
138. WHO. *International Classification of Functioning, Disability and Health: ICF.* World Health Organization; 2001.
139. Spadoni GF, Stratford PW, Solomon PE, Wishart LR. The evaluation of change in pain intensity: a comparison of the P4 and single-item numeric pain rating scales. *Journal of orthopaedic & sports physical therapy.* 2004;34(4):187-193.
140. Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. *Advances in Health Sciences Education.* 2015;20(3):829-834.
141. Widanarko B, Legg S, Devereux J, Stevenson M. The combined effect of physical, psychosocial/organisational and/or environmental risk factors on the presence of work-related musculoskeletal symptoms and its consequences. *Applied ergonomics.* 2014;45(6):1610-1621.
142. Kraatz S, Lang J, Kraus T, Münster E, Ochsmann E. The incremental effect of psychosocial workplace factors on the development of neck and shoulder disorders: a systematic review of longitudinal studies. *International archives of occupational and environmental health.* 2013;86(4):375-395.
143. Miaskowski C, Paul SM, Cooper B, et al. Identification of patient subgroups and risk factors for persistent arm/shoulder pain following breast cancer surgery. *European Journal of Oncology Nursing.* 2014;18(3):242-253.
144. Chester R, Jerosch-Herold C, Lewis J, Shepstone L. Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: a

www.manaraa.com

multicentre longitudinal cohort study. *Br J Sports Med.* 2016:bjsports-2016-096084.
145. Puentedura EJ, Cleland JA, Landers MR, Mintken PE, Louw A, Fernandez-de-Las-Penas C. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from thrust joint manipulation to the cervical spine. *J Orthop Sports Phys Ther.* 2012;42(7):577-592.
146. Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Annals of internal medicine.* 2004;141(12):920-928.
147. Bialosky JE, Bishop MD, Cleland JA. Individual expectation: an overlooked, but pertinent, factor in the treatment of individuals experiencing musculoskeletal pain. *Physical therapy.* 2016;90(9):1345-1355.
148. FOTO Completion Data. In: St. Luke's University Health Network; 2017.
149. Hart DL, Deutscher D, Werneke MW, Holder J, Wang YC. Implementing computerized adaptive tests in routine clinical practice: experience implementing CATs. *Journal of applied measurement.* 2010;11(3):288-303.
150. Gozalo PL, Resnik LJ, Silver B. Benchmarking Outpatient Rehabilitation Clinics Using Functional Status Outcomes. *Health Services Research.* 2015.

www.manaraa.com