

PATIENT-REPORTED OUTCOMES IN A PHYSICALLY ACTIVE POPULATION

A Thesis  
Submitted to the Graduate Faculty  
of the  
North Dakota State University  
of Agriculture and Applied Science

By

Christopher Ryan Peroutka

In Partial Fulfillment of the Requirements  
for the Degree of  
MASTER OF SCIENCE

Major Department:  
Health, Nutrition, and Exercise Science

March 2018

Fargo, North Dakota

North Dakota State University  
Graduate School

---

**Title**

Patient-Reported Outcomes in a Physically Active Population

**By**

Christopher Ryan Peroutka

The Supervisory Committee certifies that this *disquisition* complies with North Dakota State University's regulations and meets the accepted standards for the degree of

**MASTER OF SCIENCE**

SUPERVISORY COMMITTEE:

Shannon David

Chair

Benjamin Balas

Jim Deal

---

Approved:

March 29, 2018

Date

Yeong Rhee

Department Chair

## ABSTRACT

**OBJECTIVE:** To examine the efficacy of patient-reported outcome measures in the physically active and to determine whether a relationship exists between general health-related quality of life and specific outcome measures. Instruments used were the DPA, the SF-36, the QuickDASH, and the ASES. **METHODS:** 42 NCAA Division I athletes completed outcome measures three separate times. Repeated-measures ANOVA was performed and bivariate Pearson correlations were calculated. Additionally, test-retest reliability and minimal detectable change were assessed. Significance was set at  $\alpha \leq 0.05$ . **RESULTS:** Significant relationships were found between the DPA and several subscales of the SF-36. Participation in physical activity did not have an effect on scores as measured at different time points. **CONCLUSIONS:** The DPA and SF-36 are effective measures of health in physically active populations. The QuickDASH and ASES may not be reliable measures in these populations.

## **ACKNOWLEDGEMENTS**

Thank you to North Dakota State University Athletics for allowing the use of its facilities during data collection. The Masters of Science in Advanced Athletic Training program, North Dakota State University, provided funding for this study. I would also like to thank Dr. James Deal and Dr. Benjamin Balas for taking time out of their schedules to serve on my committee. Their insight was tremendously beneficial to my research. Finally, a special thank you to Dr. Shannon David, without whom I would have been truly lost in this endeavor. Her guidance at every step was instrumental to finishing the product that you see here today.

## TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	ix
LIST OF ABBREVIATIONS.....	x
CHAPTER I. INTRODUCTION.....	1
Purpose.....	2
Research Questions.....	2
Inclusion Criteria.....	2
Exclusion Criteria.....	3
Limitations.....	3
Delimitations.....	3
Definition of Terms.....	3
CHAPTER II. REVIEW OF LITERATURE.....	5
Introduction.....	5
Patient-Oriented Care.....	5
Types of PRO Measures.....	6
Considerations when Choosing a PRO.....	7
Psychometric Properties of PROs.....	8
Reliability.....	9
Validity.....	9
Responsiveness.....	10
Barriers to Use.....	10
Short-Form 36-Item Health Survey.....	11
Scoring.....	12

Normative Scores .....	12
Validity .....	13
Reliability .....	13
Disablement in the Physically Active .....	14
Scoring.....	14
Normative Scores .....	15
Validity .....	15
Reliability .....	15
Responsiveness.....	15
Disabilities of the Arm, Shoulder, and Hand .....	16
Scoring.....	16
Normative Scores .....	17
Validity .....	17
Reliability .....	18
Responsiveness.....	18
QuickDASH .....	19
American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form.....	19
Scoring.....	20
Normative Scores .....	20
Validity .....	21
Reliability .....	21
Responsiveness.....	22
Summary .....	22
CHAPTER III. METHODOLOGY .....	24
Purpose .....	24

Research Design .....	24
Sample of Study .....	24
Instrumentation.....	25
Procedure.....	26
Statistical Analysis .....	27
CHAPTER IV. MANUSCRIPT .....	29
Introduction .....	29
Methodology .....	30
Participants .....	30
Procedures .....	31
Instrumentation.....	32
Disablement in the Physically Active Scale .....	32
Short-Form 36-Item Health Survey .....	32
Shortened Disabilities of the Arm, Shoulder, and Hand.....	33
American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form .....	33
Statistical Analysis .....	33
Results .....	34
Relationships between PRO Scores.....	34
Appropriateness of PROs in Physically Active Populations.....	35
Effect of Administration Time Point on PRO Scores.....	35
Test-Retest Reliability and Minimal Detectable Change.....	36
Discussion .....	37
Relationships between PRO Scores.....	37
Appropriateness of PROs in Physically Active Populations.....	38
Effect of Administration Time Point on PRO Scores.....	38

Test-Retest Reliability and MDC.....	39
Limitations.....	40
Conclusions .....	40
REFERENCES .....	42
APPENDIX A. DEMOGRAPHIC QUESTIONNAIRE .....	46
APPENDIX B. ORAL SCRIPT FOR RESEARCHER.....	47



## LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Mean (SD) Normative Values for the SF-36 as Established in Previous Studies .....	13
2. Internal Consistency and Test-Retest Reliability for the Dimensions of the SF-36.....	14
3. Demographic Information.....	31
4. Patient-Reported Outcome Measure Scores (n = 42) .....	34
5. Relationships between Patient-Reported Outcome Measure Scores (n = 42) .....	35
6. PRO Scores and Repeated-Measures ANOVA (n = 35) .....	36
7. Test-Retest Reliability and Minimal Detectable Change (n = 35).....	36

## LIST OF ABBREVIATIONS

PRO.....	Patient-Reported Outcome
HRQOL.....	Health-Related Quality of Life
SF-36.....	Short-Form 36-Item Health Survey
SF-12.....	Short-Form 12-Item Health Survey
DPA.....	Disablement in the Physically Active
DASH.....	Disabilities of the Arm, Shoulder, and Hand
ASES.....	American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form
p-ASES .....	ASES Patient Self-Assessment Section
NDSU.....	North Dakota State University
NCAA .....	National Collegiate Athletics Association
POEM .....	Patient-Oriented Evidence that Matters
SANE .....	Single Assessment Numeric Evaluation
EQ-5D.....	EuroQol Five Dimensions Questionnaire
IKDC.....	International Knee Documentation Committee Subjective Knee Form
ICC .....	Intraclass Correlation Coefficient
SRM .....	Standardized Response Mean
ES .....	Effect Size
MDC .....	Minimal Detectable Change
MCID .....	Minimal Clinically Important Difference
MOS.....	Medical Outcomes Study
GF .....	Global Functioning
ADL .....	Activities of Daily Living
VAS.....	Visual Analog Scale

## CHAPTER I. INTRODUCTION

Evidence-based practice (EBP) is the foundation for health care administration in the sports medicine setting. EBP is the integration of a clinician's personal experience with the best available research and the values of their patients.<sup>1</sup> Clinical outcomes assessment enables EBP through the examination of patient values after medical interventions. It provides clinicians with means to assess treatment progress and measure the end results of services rendered.<sup>2</sup> In short, clinical outcomes assessment provides a means for measuring and tracking health-related quality of life.

Health-related quality of life (HRQOL) is an individual's general feeling of overall well-being.<sup>3</sup> HRQOL encompasses physical, psychological, and social aspects of health as derived from personal beliefs, preferences, experiences, and expectations.<sup>3-8</sup> Parsons and Snyder<sup>3</sup> note that HRQOL is both individualistic and variable in its nature; it is influenced by injury, illness, and disease, as well as previous personal experiences and changing values and priorities. One way to measure HRQOL is by using patient-reported outcome (PRO) measures. To be effective, PROs must be able to fully capture the impact of injury and response to treatment on individual patients.<sup>1</sup>

Patient-reported outcome measures are self-report questionnaires on a patient's health condition and include aspects such as symptoms, physical function, and general well-being.<sup>2,9</sup> PROs are either generic or specific in nature. Generic instruments measure overall wellbeing in a wide variety of patients independent of injury or condition whereas specific measures evaluate aspects related to a precise disease, injury, population, or anatomical region.<sup>2,3,8-10</sup> The target populations for outcome measures are typically as broad as possible, but often do not take into account that HRQOL varies between physically active and sedentary individuals. Many

instruments are not valid in the physically active, or have not been tested in these populations which may limit their use in a sports medicine setting.<sup>1,5,11-14</sup>

Many patient-reported outcome measures have not been fully evaluated in physically active individuals. Studies have shown that quality of life differs in physically active individuals compared to the general population.<sup>11-15</sup> Many of these studies<sup>11,13,14</sup> assess HRQOL using the SF-36 (Short-Form 36-Item Health Survey), a popular generic PRO. Fewer studies exist to analyze the results for physically active individuals on specific PRO instruments such as the DPA (Disablement in the Physically Active Scale), QuickDASH (shortened Disabilities of the Arm, Shoulder, and Hand), and ASES (American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form).

### **Purpose**

The purpose of this study was to examine the effectiveness of patient-reported outcome measures in the physically active and to determine whether a relationship exists between general health-related quality of life and shoulder-specific outcome measures.

### **Research Questions**

- (1) What is the relationship between scores on shoulder-specific and scores on general health-related quality of life patient-reported outcome measures?
- (2) Are patient-reported outcome measures, specifically the SF-36, DPA, QuickDASH, and ASES, appropriate in physically active populations?

### **Inclusion Criteria**

- (1) Participants were 18 years of age or older at the time of data collection
- (2) Participants were athletes on an NDSU sports team roster at the time of data collection

- (3) Participants must have been able to read and write in English to avoid confusion with items on the outcome measures

### **Exclusion Criteria**

- (1) Individuals who were injured at the time of initial data collection or who had a history of injury within the previous thirty days

### **Limitations**

- (1) Data was obtained from a sample of collegiate athletes and may not be fully representative of the spectrum of physically active individuals
- (2) Participants may not be entirely honest when answering items on the instruments, and this cannot be controlled for due to the self-report nature of PROs
- (3) Participants were obtained from both in-season and offseason sports; little research has been completed on the effect that a sports season has on scores of the instruments used in this study

### **Delimitations**

- (1) Participants were all athletes at an NCAA Division I institution
- (2) Participants attended North Dakota State University in Fargo, North Dakota

### **Definition of Terms**

Patient-reported outcome (PRO) measures are self-report questionnaires used by clinicians to assess the patient perspective on their injury or condition and its effects on different aspects of their life.<sup>2,5,7,9,14,16</sup>

Health-related quality of life (HRQOL) is an individual's perception of their function in everyday life. This includes physical, psychological, and social aspects of health derived from personal beliefs, preferences, experiences, and expectations.<sup>3-6,8</sup>

Generic PROs measure overall wellbeing of a patient, regardless of injury or condition.<sup>2,3,8-10</sup>

Region-specific PROs evaluate aspects of health and function that are related to a particular anatomical region.<sup>2,3,10</sup>

Items are the specific questions when used in the context of PROs.<sup>9</sup>

Domains are the sub-categories that items are grouped into and provide information on specific types of health measurement.<sup>9</sup>

Psychometric properties of a PRO determine its ability to accurately and consistently measure what it intends to and include reliability, validity, and responsiveness.<sup>16</sup>

A ceiling effect is the level above which variance in an independent variable is no longer being measured.<sup>1,17</sup>

## **CHAPTER II. REVIEW OF LITERATURE**

### **Introduction**

The purpose of this study was to examine the effectiveness of patient-reported outcome measures in the physically active and to determine whether a relationship exists between general health-related quality of life and shoulder-specific outcome measures. The following research questions guided this study: (1) What is the relationship between scores on shoulder-specific and scores on general health-related quality of life patient-reported outcome measures? (2) Are patient-reported outcome measures appropriate in physically active populations?

There has been an increased emphasis on patient-oriented outcomes assessment in health care over the past decade, however many athletic trainers have been reluctant to implement them in their practice for various reasons. Patient-reported outcomes can provide athletic trainers with a mechanism for assessing the impact of a disorder or disease on a patient, evaluating treatment progress, and measuring the end results of services provided.<sup>2,10</sup> Information from this study could be useful for clinicians in sports medicine settings who want to use acceptable outcome measures for their patient demographic.

This literature review is organized into the following areas: Patient-Oriented Care, The Short-Form Health Survey, The Disablement in the Physically Active Scale, Disabilities of the Arm, Shoulder, and Hand, American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, and Summary.

### **Patient-Oriented Care**

Patient-oriented care is presently at the forefront of health care provision and research. Increased recognition of patient perspective on their functioning and health has led to an emphasis in research to develop concepts and instruments to measure health-related quality of

life (HRQOL).<sup>4</sup> HRQOL is an individual's perceived function in everyday life and includes physical, psychological, and social aspects of health derived from personal beliefs, preferences, experiences, and expectations.<sup>3-6,8</sup> HRQOL is a form of patient-oriented evidence that matters (POEM), an area which consists of items that matter to patients: symptom improvement, morbidity, financial considerations, and quality of life.<sup>3</sup>

Patient-reported outcome (PRO) measures are self-report questionnaires used by clinicians to assess the patient perspective on their injury or condition and its effects on different aspects of their life. A wide range of health status factors are taken into account with PROs including symptoms, physical function, and general well-being, among others.<sup>2,16</sup> PRO instruments typically contain both objective and subjective domains with the plan to completely capture evidence that truly matters to the patient. An objective assessment of functioning is important in defining a patient's degree of health, but the patient's subjective perceptions and expectations translate the objective assessment into actual quality of life experienced.<sup>8</sup> Patient-reported outcome instruments enable clinicians to measure and track HRQOL in their patients by describing the burden of a disease, injury, or condition and its effects on the patient's ability to participate in activity or perform normal, everyday functions.<sup>4</sup>

### **Types of PRO Measures**

PRO instruments can be generally divided into generic or specific measures. Generic PROs measure overall wellbeing in a wide variety of patients, regardless of injury or condition.<sup>2,3,8-10</sup> These measures cover a broad range of health status at the expense of item depth and responsiveness to change.<sup>2,3,8</sup> The benefit of such a measure is that it can be applied without regards to specific injury conditions or bodily region in many different patients. Examples of generic PROs include the MOS 36-Item Short Form Health Survey (SF-36) and its 12 item



shortened version (SF-12), the Single Assessment Numeric Evaluation (SANE), and the EuroQol Five Dimensions questionnaire (EQ-5D). PROs that will be discussed in this literature review are the SF-36 and SF-12.

Specific measures evaluate aspects related to a precise disease, injury, population, or anatomical region.<sup>2</sup> While other types of specific PROs have their place in general health administration, region-specific PROs are more relevant in sports medicine settings because they evaluate aspects of health and function that are related to a particular anatomical region.<sup>2,3,10</sup> The breadth of the regions covered by these PROs can vary; some evaluate only a specific joint or body part while others evaluate entire limbs. Examples of region-specific PROs include the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH) and its shortened version (QuickDASH), the American Shoulder and Elbow Surgeons shoulder assessment (ASES), the Oswestry Low Back Pain Disability Questionnaire, and the International Knee Documentation Committee Subjective Knee Form (IKDC). The DASH, QuickDASH, and ASES will be further discussed in this literature review. Additionally, the Disablement in the Physically Active Scale (DPA) is an example of a population-specific PRO for use on physically active populations, and it will be further discussed in this literature review.

### **Considerations when Choosing a PRO**

It is important for clinicians and researchers to understand the variety of PRO instruments that exist and the many aspects that must be taken into consideration when selecting one to use in clinical practice. The decision should ensure that the instrument is psychometrically sound and aligns with its intended purpose, and the instrument should be relevant to the target population as well as the injury or condition being assessed.<sup>3,8,18</sup> A concern is that many current outcome

measures have limited applicability in athletic populations, so research is necessary to determine the appropriateness of an instrument for use on competitive athletes.<sup>2</sup>

Another important consideration is the burden of the PRO on the patient and the clinician.<sup>3,18</sup> Distress to the patient can be minimized by using instruments that take less time to complete and are succinctly worded.<sup>3</sup> For example, the SF-36 contains 36 items, which may take considerably longer to complete than the 12-item SF-12. Using the SF-12 may decrease the time burden on the patient, but it may do so at the expense of reliability and validity. The burden of the instrument on clinician administration should also be considered.<sup>18</sup> This may include training involved, costs for obtaining and using the instrument, and time required for administration of, scoring, and analyzing the results.<sup>3,18</sup> A clinician may be more likely to use a PRO instrument if there is a small burden on both the patient and the clinician.

When choosing an instrument, one must also consider its accessibility and whether it is copyright protected.<sup>18</sup> Despite the fact that PROs are only simple questionnaires, many require purchase so they can be properly administered and scored. Some instruments, such as the SF-36, may even provide software that will score the responses automatically. Instruments with complicated scoring algorithms may decrease clinician burden by making it easier to evaluate and interpret the results.

### **Psychometric Properties of PROs**

The ability of a PRO to accurately and consistently measure intended outcomes is determined by its psychometric properties.<sup>16</sup> A few aspects of evidence that can be used to determine the psychometric properties of a PRO include reliability, validity, and responsiveness.

### ***Reliability***

Reliability refers to the overall consistency of a measure, or its ability to yield the same results each time it is administered.<sup>8,16,18,19</sup> Subsets of reliability include test-retest reliability, internal consistency reliability, inter-rater reliability, and parallel-forms reliability. Test-retest reliability is the repeatability of the measure across multiple time points while other variables remain unchanged; it is measured using intra-class correlation coefficient (ICC).<sup>16,19</sup> Internal consistency reliability is how well items measure the same domain and is measured using Cronbach's  $\alpha$ .<sup>16,18,19</sup> ICC and Cronbach's  $\alpha$  values range between zero and one, with 0.70 the standard threshold for adequate reliability.<sup>16</sup> Inter-rater reliability is the degree to which raters agree with each other.<sup>19</sup> Parallel-forms reliability is the degree to which different forms of the same measure agree with each other when administered to the same individuals. Assessments of reliability in PROs typically report test-retest and internal consistency reliability.<sup>19</sup>

### ***Validity***

Validity is the extent to which an instrument measures what it was intended to measure and can be divided into internal and external validity.<sup>1,8,18,19</sup> Internal validity is the ability of an instrument to control for possible effects of extraneous variables on what is being measured while external validity is the ability to generalize the results to other groups beyond those in the experiment. Internal validity can be further divided into content, criterion, and construct validity. Content validity is the degree to which the measure represents all facets of a construct.<sup>18,19</sup> Criterion validity is the degree to which scores on the measure correlate with other measures and includes concurrent (degree of correlation to a gold standard) and predictive (degree to which scores represent future behavior) validity.<sup>1,18,19</sup> Finally, construct validity is the degree to which a measure captures what it claims to measure and includes convergent (correlation with similar

constructs) and discriminant (lack of correlation with dissimilar constructs) validity.<sup>1,18,19</sup>

Pearson correlation coefficient ( $r$ ) is a measure of dependence between two variables and is often used in types of validity that require correlations. It ranges from negative one to positive one; the strength of correlation increases as it gets further away from zero.

### ***Responsiveness***

Responsiveness is an aspect of validity that measures an instrument's ability to capture change.<sup>8,16,18,20</sup> PROs may be administered at different time points in order to detect change in an injury or condition over time. Instruments should be able to determine when a change in the score is clinically significant. Measures of responsiveness include standardized response mean (SRM) and effect size (ES) while minimal detectable change (MDC) and minimal clinically important different (MCID) are two measures of the significance of change. MDC is the smallest amount of change that needs to occur in a score to demonstrate true change beyond error.<sup>5,21</sup> MCID is the smallest change in an outcome that a patient would identify as important.<sup>20,21</sup> Turner et al.<sup>21</sup> argued that MCID is more reliable than MDC given that it is calculated using an anchor, an external criterion to interpret whether a particular magnitude of change is significant.<sup>21</sup> When anchor-based methods are unavailable, the MDC can be used as a replacement for MCID.<sup>21</sup>

### **Barriers to Use**

The use of PROs in athletic training does not come without drawbacks, and subsequently barriers to their use. One of the common barriers cited in the literature is the lack of time associated with completing, scoring, and interpreting the results of instruments.<sup>9,10,22</sup> This may be due to large patient load of the clinician or other time-consuming responsibilities such as practice and game coverage.

Another common barrier is the lack of measures appropriate for the athletic training setting as well as low organizational support to use PROs in practice.<sup>2,9,22</sup> There is hesitance to implement mandated PRO use in athletic training because there are few acceptable measures available. Clinicians also show concern about patients' ability to complete outcome measures due to questions which may be too complicated or confusing.<sup>3,10,22</sup> Organizations could potentially improve outcomes assessment by providing the necessary support in terms of training and financial resources available to the clinicians.

A limitation to PROs is the possibility that ceiling effects exist associated with their use on physically active individuals. Most instruments were not constructed to measure high functional ability seen in the physically active.<sup>1</sup> In statistics, a ceiling effect is the level above which variance in an independent variable is no longer measured.<sup>1,17</sup> This may be of concern with some PROs used in a physically active population; an athlete may have a normal score on a PRO before reaching full baseline physical function.<sup>1,17</sup> Ceiling effects may limit the validity of instruments when used in a physically active population which in turn limits their usefulness in a sports medicine clinical setting.

### **Short-Form 36-Item Health Survey**

The Short-Form 36-Item Health Survey (SF-36) is a widely used generic PRO designed for use in a wide variety of populations. It was developed for the Medical Outcomes Study (MOS), a large-scale study of how patients fared with health care in the United States, and published in 1992.<sup>23-25</sup> The SF-36 measures functional status, wellbeing, and overall evaluation of health in its eight dimensions.<sup>26</sup> These dimensions include Physical Functioning, Social Functioning, Role Limitations due to Physical Problems, Bodily Pain, General Mental Health, Role Limitations due to Emotional Problems, Vitality, and General Health Perceptions.<sup>24,26</sup> A

second version of the SF-36 (version 2.0) was released in 1996 to improve scales and simplify instructions.<sup>25</sup> Both versions are available for use today.

### **Scoring**

Each dimension of the SF-36 is scored and reported separately so that eight scores are produced upon completion of the questionnaire.<sup>25</sup> Item responses are in the form of two-, three-, five-, and six-point Likert scales. Each response is then recoded into a value out of 100 and the statistical mean for items of each scale is taken. For example, question one can receive a score of one, two, three, four, or five. These scores would be recoded into values of 100, 75, 50, 25, or zero respectively, and then averaged with the scores of the other items in the same scale. Each of the eight scales produces a value ranging from zero to 100, where higher scores represent better health states.<sup>23</sup> Two summary measures can be derived from the eight dimensions of the SF-36 if desired: physical and mental component scores.

### **Normative Scores**

Normative scores for each of the dimensions of the SF-36 have been established in various studies.<sup>11,27,28</sup> These values can be found in Table 1. Jenkinson et al.<sup>28</sup> reported normative scores broken down by age, gender, and occupation. Values reported in Table 1 are of males and females aged 18 to 24. McAllister et al.<sup>13</sup> conducted a study to establish baseline SF-36 data for uninjured Division I collegiate athletes. Huffman et al.<sup>11</sup> compared normative values in Division I and II National Collegiate Athletic Association (NCAA) athletes to an age-matched general population. Values for the general population were obtained from the Medical Outcomes Study (MOS).

**Table 1. Mean (SD) Normative Values for the SF-36 as Established in Previous Studies**

	Jenkinson et al. <sup>a</sup>		McAllister et al. <sup>b</sup>		Huffman et al. <sup>c</sup>	
	Male 18-24	Female 18-24	Male	Female	NCAA	MOS
<b>PF<sup>d</sup></b>	92.8 (16.8)	90.1 (16.4)	94 (1.2)	96 (1.2)	97.7 (8.0)	92.1 (18.3)
<b>SF</b>	90.2 (16.4)	85.7 (19.7)	88 (1.2)	87 (1.6)	94.5 (11.9)	83.9 (20.6)
<b>RLP</b>	91.8 (22.6)	88.6 (25.5)	96 (1.0)	91 (2.0)	94.5 (17.8)	89.1 (26.8)
<b>RLE</b>	82.9 (31.1)	78.8 (33.0)	94 (1.2)	93 (1.8)	82.6 (10.7)	83.0 (31.1)
<b>MH</b>	74.8 (15.4)	70.2 (17.4)	80 (1.0)	79 (1.0)	82.6 (10.7)	74.7 (18.1)
<b>Vitality</b>	66.4 (17.1)	59.8 (19.4)	69 (1.1)	68 (1.1)	69.1 (13.7)	62.5 (19.8)
<b>Pain</b>	86.6 (17.9)	81.7 (20.8)	84 (1.1)	82 (1.4)	83.9 (17.9)	80.8 (21.3)
<b>GHP</b>	72.0 (20.1)	72.1 (20.3)	81 (1.1)	79 (1.3)	84.5 (13.4)	76.7 (18.2)

<sup>a</sup>Data from Jenkinson et al.<sup>28</sup><sup>b</sup>Data from McAllister et al.<sup>13</sup><sup>c</sup>Data from Huffman et al.<sup>11</sup><sup>d</sup>Abbreviations: PF = Physical functioning; SF = Social functioning; RLP = Physical role limitations; RLE = Emotional role limitations; MH = Mental health; GHP = General health perceptions; NCAA = National Collegiate Athletic Association; MOS = Medical Outcomes Survey

## Validity

Validity of the SF-36 has been explored in a few studies.<sup>26,29,30</sup> Using information from the Medical Outcomes Survey, McHorney et al.<sup>30</sup> determined that convergent validity was adequate in every dimension except Global Health Perception which was rated as poor. This study also found that discriminant validity was adequate in every dimension excluding Social Functioning and Vitality. McHorney et al.<sup>30</sup> claimed that the Physical Function dimension was the most valid measure of aspects of physical health while the Mental Health dimension was the most valid measure of aspects of mental health. Brazier et al.<sup>26</sup> reported convergent and discriminant validity of the overall SF-36 and Jenkinson et al.<sup>29</sup> reported criterion validity of the SF-36 when compared to patient reports of overall general health.

## Reliability

Reliability of the eight dimensions of the SF-36 has been assessed in a number of studies.<sup>26,29,31</sup> McHorney et al.<sup>31</sup> and Jenkinson et al.<sup>29</sup> reported Cronbach  $\alpha$  levels for internal consistency reliability of the domains in their respective studies while Brazier et al.<sup>26</sup> reported Cronbach  $\alpha$  and ICC values for internal consistency and test-retest reliability respectively. These

values can be found in Table 2. Each of the authors reported that internal consistency reliability was at least adequate for each of the eight dimensions of the SF-36.<sup>26,29,31</sup> Brazier et al.<sup>26</sup> indicated that test-retest reliability of the SF-36 was excellent.

**Table 2. Internal Consistency and Test-Retest Reliability for the Dimensions of the SF-36**

	McHorney et al. <sup>a</sup>	Jenkinson et al. <sup>b</sup>	Brazier et al. <sup>c</sup>	
	Cronbach $\alpha$	Cronbach $\alpha$	Cronbach $\alpha$	ICC
<b>PF<sup>d</sup></b>	.49-.80	.90	.93	.81
<b>SF</b>	.74	.76	.73	.60
<b>RLP</b>	.65-.70	.88	.96	.69
<b>RLE</b>	.63-.73	.80	.96	.63
<b>MH</b>	.65-.81	.83	.95	.75
<b>Vitality</b>	.69-.81	.85	.96	.80
<b>Pain</b>	.70	.82	.85	.78
<b>GHP</b>	.38-.72	No Data	.95	.80

<sup>a</sup>Data from McHorney et al.<sup>31</sup>

<sup>b</sup>Data from Jenkinson et al.<sup>29</sup>

<sup>c</sup>Data from Brazier et al.<sup>26</sup>

<sup>d</sup>Abbreviations: PF = Physical functioning; SF = Social functioning; RLP = Physical role limitations; RLE = Emotional role limitations; MH = Mental health; GHP = General health perceptions; ICC = Intraclass correlation coefficient

### Disablement in the Physically Active

The Disablement in the Physically Active scale (DPA) is a population-specific PRO designed to assess HRQOL in physically active populations.<sup>5,7,32</sup> It is a 16-item questionnaire that covers four domains based on themes found in the Institute of Medicine disablement model: impairment, functional limitation, disability, and quality of life.<sup>5,32,33</sup> The DPA was developed by Vela and Denegar<sup>32,33</sup> and released in 2010.

### Scoring

Items on the DPA are scored using a five-point Likert scale, where a one indicates *no problem* and a five indicates *severe problem*.<sup>33</sup> A final score is obtained by summing the scores from the 16 items and subtracting 16 points. DPA scores will range from zero to 64, with a higher score signifying a higher level of disablement.<sup>5,32</sup>



## **Normative Scores**

Vela and Denegar<sup>32</sup> reported standard values for the DPA in healthy, physically active adolescents and young adults. Participants were male and female competitive and recreational athletes (mean age  $20.1 \pm 3.8$  years). Mean scores were found to be  $3.68 \pm 5.65$  in this sample. The authors noted that no ceiling effects occurred.<sup>32</sup>

## **Validity**

The developers of the DPA performed a psychometric assessment on their outcome measure to assess its validity, reliability, and responsiveness in young, physically active participants.<sup>32,33</sup> Concurrent validity was assessed by comparing scores on the DPA to those on the Global Functioning (GF) scale, which is considered the gold standard to establish concurrent validity.<sup>32</sup> There was a significant correlation between scores on the DPA and the GF for acute ( $r = -0.751, P < .001$ ) and chronic ( $r = -0.714, P < .001$ ) injuries indicating concurrent validity of the DPA.<sup>32</sup> Additional measurements of validity were not available in the literature.

## **Reliability**

Reliability of the DPA was assessed using intraclass correlation coefficient (ICC) for test-retest reliability and Cronbach  $\alpha$  for internal consistency. It was determined that test-retest reliability was excellent (ICC = .943) while internal consistency was adequate for both acute (Cronbach  $\alpha = .908$ ) and chronic (Cronbach  $\alpha = .890$ ) injuries.<sup>32</sup> Test-retest reliability of the DPA was assessed in a study conducted by Hoch et al.<sup>5</sup> which reported an ICC of .792 in a sample of 16 female collegiate soccer players.

## **Responsiveness**

Responsiveness of the DPA was established by Vela et al.<sup>32</sup> using minimal clinically important difference (MCID). An MCID value of nine points was established for acute injuries

while an MCID value of six points was established for chronic injuries.<sup>32</sup> A patient has experienced a clinically significant change in their condition if their DPA score is greater than the respective MCID. Hoch et al.<sup>5</sup> measured the minimal detectable change (MDC) of the DPA to be 12.48, which was greater than the MCID value established by Vela et al.<sup>32</sup> These results suggest that a score change of greater than 12 points may be required to signify true change in a patient's health status when using the DPA.

### **Disabilities of the Arm, Shoulder, and Hand**

The Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire is a 30-item region-specific outcome measure for the upper extremity. Originally released in 1996, the DASH has become a widely used measure of symptoms and physical function related to disability in the upper limb.<sup>34-38</sup> The DASH was designed to be used on a wide variety of patients with one or more disabilities as a means to describe the disability and monitor changes over time.<sup>34,38</sup> It assesses six domains which include daily activities, symptoms, social function, work function, sleep, and confidence.<sup>36</sup>

### **Scoring**

The DASH questionnaire takes approximately five minutes to complete with each of the 30 items scored on a five-point Likert scale, where one signifies *no difficulty/ symptoms* and five signifies *extreme difficulty/ symptoms*.<sup>34,36-39</sup> The total score is obtained by summing the responses, subtracting 30, and dividing by 1.2 so that the final score falls between zero and 100.<sup>34,35</sup> Higher scores reflect greater disability. The responder can miss up to three items at which point the mean values of their other responses would be substituted for each missing item.<sup>35-37,39</sup> There are four additional items for each of work and sports/ performing arts that are

typically used for workers and athletes respectively; these are scored and reported separately using the same zero to 100 scale.<sup>37,39</sup>

### **Normative Scores**

Several studies have measured normative scores on the DASH, scores a clinician would expect to see from healthy individuals.<sup>17,36,39-41</sup> Hunsaker et al.<sup>41</sup> suggest a mean score of  $10.10 \pm 14.68$  while Clarke et al.<sup>40</sup> suggest a mean of  $1.85 \pm 5.99$ . Hsu et al.<sup>17</sup> measured normative scores in collegiate athletes and found an overall mean of  $1.37 \pm 2.96$ . This study also considered the variables of gender, participation in an overhead sport, and prior upper extremity injury. The results suggest that there is a significant effect of gender (male mean  $0.98 \pm 2.60$  compared to female mean  $1.82 \pm 3.27$ ;  $P = .010$ ) and participation in an overhead sport (overhead athlete mean  $1.81 \pm 3.57$  compared to non-overhead athlete mean  $0.98 \pm 2.60$ ;  $P = .04$ ) on normative DASH scores in collegiate athletes.<sup>17</sup> It is important to understand that differences exist between athletes and non-athletes as well as between athletes of different sports in terms of scores on the DASH. There may be a significant ceiling effect of the DASH in athletes which is indicated when the best possible score is achieved by 15% to 20% of patients.<sup>17</sup>

### **Validity**

Validity of the DASH was assessed in various studies and reviews.<sup>35,39,42-44</sup> Content, discriminant, and construct validity of the DASH were established in various studies<sup>35,42,44</sup> on the following populations: patients with various upper limb conditions, patients with post-surgical upper limb conditions, and patients with proximal humeral fractures. Michener and Leggin<sup>43</sup> conducted a review and reported that the DASH displayed content, construct, and discriminant validity in patients with musculoskeletal disorders. Hsu et al.<sup>17</sup> suggest that the validity of the DASH in athletes may be limited by a ceiling effect which means that a score of zero may not

represent normal physical function in athletes. This could be a limitation to use of the DASH in the physically active.

### **Reliability**

Test-retest and internal consistency reliability of the DASH were demonstrated in a number of studies.<sup>35,37,39,42,43</sup> Test-retest reliability was measured by Beaton et al.<sup>35</sup> who reported an ICC of .96 while Michener and Leggin<sup>43</sup> reported an ICC of .92 from their review. Angst et al.<sup>39</sup> found that ICC varied between .93 and .95 in their systematic review. Internal consistency reliability was reported at a Cronbach  $\alpha$  of .95 by Franchignoni et al.<sup>37</sup> while Gummesson et al.<sup>42</sup> reported Cronbach  $\alpha$  of .97 preoperatively and .98 postoperatively in patients undergoing shoulder surgery. Michener and Leggin<sup>43</sup> reported a Cronbach  $\alpha$  of .96 and Angst et al.<sup>39</sup> reported a range from .92 to .98. These reported values indicate that the DASH has excellent test-retest and internal consistency reliability.

### **Responsiveness**

Responsiveness of the DASH was established using standardized response mean (SRM) and effect size (ES), and one study also reported MDC.<sup>35,42</sup> In an evaluation of shoulder patients who observed a change in their condition between assessments, Beaton et al.<sup>35</sup> reported an SRM of 0.81 and an ES of 0.64. This study also found the MDC to be 12.75 with a 95% confidence interval. The authors of this study concluded that the DASH is responsive to change in patients with various shoulder conditions.<sup>35</sup> Gummesson et al.<sup>42</sup> reported an SRM of 1.2 and ES of 0.7 in their study and concluded that the DASH was responsive to change in patients with upper-extremity musculoskeletal injuries undergoing surgical treatment.

## **QuickDASH**

A concern with the DASH is that it contains 30 items which is a large number for a PRO. High completion time is a burden on the patient and may result in altered responses. The QuickDASH was created in 2005 by Beaton et al.<sup>45</sup> to reduce the time burden on the respondent and the administrator by eliminating item redundancy. This shortened version of the DASH contains 11 items and is intended to measure all of the same domains. Scores are reported on the same zero to 100 scale, where higher scores signify greater disability. Test-retest reliability (ICC = 0.93) and cross-sectional reliability (Cronbach's  $\alpha = 0.92$ ) were excellent in a study conducted by Gummesson et al.<sup>46</sup> Mintken et al.<sup>47</sup> also reported excellent test-retest reliability (ICC = 0.90) of the DASH while Franchignoni et al.<sup>48</sup> reported good internal consistency (Cronbach's  $\alpha = 0.87$ ). Responsiveness of the QuickDASH was assessed by Polson et al.<sup>49</sup> and a SRM of 1.1 was measured. MDC was found to be 11.2 by Mintken et al.<sup>47</sup> and 11 by Polson et al.<sup>49</sup> The results of the various studies indicate that the QuickDASH is a valid, reliable, and responsive PRO that can be used in place of the DASH in clinical settings.<sup>46-49</sup>

## **American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form**

The American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES) is a region-specific patient-reported outcome (PRO) measure for assessing shoulder function. Released in 1994, the ASES was designed as an easy to use method of assessing activities of daily living based on patient self-evaluation.<sup>50</sup> The form consists of a patient self-assessment section and a physician assessment section; however the physician assessment section is not used for scoring purposes. The patient section can be administered without the physician section and therefore requires approximately three minutes for the respondent to complete.<sup>50</sup>

## Scoring

The patient self-assessment section of the ASES (p-ASES) is divided into three sections: pain (six items), instability (two items), and activities of daily living (ADL, 10 items for each shoulder). The pain and instability sections contain both yes/no questions and visual analog scales (VAS) which are scored from zero (best) to 10 (worst). The items in the ADL section are scored on a four-point Likert scale, where zero indicates *inability to perform* and three indicates *no difficulty*.<sup>39,50</sup> Scores are interpreted on a zero to 100 scale where higher scores are more desirable.

## Normative Scores

Normative scores for healthy individuals have been established for the p-ASES in various studies.<sup>40,51,52</sup> In a study of young, active adults performed by Clarke et al.<sup>40</sup>, a mean score of  $98.9 \pm 3.3$  was established for the p-ASES. Sallay and Reed<sup>52</sup> reported a mean score of  $92.2 \pm 14.5$  in patients from an outpatient orthopedic center being seen for conditions unrelated to the shoulder. This study also suggests that sports participation may have a positive correlation with ASES scores ( $p = 0.0007$ ), though the study did not indicate specific sports.<sup>52</sup> Brinker et al.<sup>51</sup> calculated ASES scores in healthy athletes grouped by their gender and sport type and found no significant differences between gender or between NCAA throwing athletes, NCAA non-throwing athletes, and recreational athletes. This study reported a mean of 98.5 (range 88.5 – 100) for all groups and a slightly higher mean of 98.7 (range 95 – 100) for the NCAA throwing athlete group (standard deviation was not reported).<sup>51</sup> These normative scores indicate that perfect scores on the p-ASES may not be necessary to demonstrate normal health and function in the shoulder.

## **Validity**

Validity of the ASES has been assessed in some studies.<sup>53,54</sup> Michener et al.<sup>53</sup> assessed convergent and discriminant validity in patients with various shoulder pathologies. Convergent validity was confirmed by examining the relationship between the ASES scores and those of the Penn scale ( $r = 0.78$ ;  $P < .01$ ), the SF-36 physical function score ( $r = 0.41$ ;  $P = .001$ ), and the SF-36 physical component summary score ( $r = 0.40$ ;  $P = .001$ ). Discriminant validity was confirmed by comparing ASES scores of patients with varying levels of condition improvement. Oh et al.<sup>54</sup> reported construct validity of the ASES when comparing scores with the Simple Shoulder Test ( $r = 0.350$ ;  $P < .01$ ), Constant Score ( $r = 0.356$ ;  $P < .01$ ), and UCLA shoulder score ( $r = 0.373$ ;  $P < .05$ ). Criterion validity was also reported by Oh et al.<sup>54</sup> when using the SF-36 as a standard for comparison. Significant Pearson correlation coefficients were found between the ASES and SF-36 dimensions Physical Function ( $r = 0.266$ ;  $P < .01$ ), Role Physical ( $r = 0.208$ ;  $P < .01$ ), Social Functioning ( $r = 0.179$ ;  $P < .05$ ), and Physical Component Score ( $r = 0.199$ ;  $P < .01$ ).

## **Reliability**

Internal consistency and test-retest reliability were assessed in various studies.<sup>39,53-55</sup> Internal consistency reliability was established by Michener et al.<sup>53</sup> at a Cronbach's  $\alpha$  level of 0.86 while Oh et al.<sup>54</sup> calculated the internal consistency of the different sections of the ASES and found Cronbach's  $\alpha$  values of .711 for pain, .850 for ADL, and .970 for instability. Michener et al.<sup>53</sup> calculated test-retest reliability with an ICC of 0.84 while Beaton et al.<sup>55</sup> reported an ICC of 0.96. In their systematic review, Angst et al.<sup>39</sup> found Cronbach's  $\alpha$  values of .61 to .96 for internal consistency reliability and ICC values of .84 to .96 for test-retest reliability. Cronbach's  $\alpha$  and ICC values greater than 0.7 indicate adequate reliability, therefore the ASES is a reliable measure of shoulder function.<sup>16,53-55</sup>

## **Responsiveness**

Responsiveness of the ASES was reported in various studies using SRM and ES while some MDC values were also reported.<sup>53-55</sup> Michener et al.<sup>53</sup> calculated an SRM of 1.54 and an ES of 1.39 while MDC was reportedly 9.4 at the 90% confidence interval. Oh et al.<sup>54</sup> reported an SRM of .771 and ES of .617 while Beaton et al.<sup>55</sup> found an SRM of .93. In their systematic review, Angst et al.<sup>39</sup> found that SRM ranged from 1.42 to 1.81 and ES ranged from .93 to 3.53. An MDC of 11.2 was also reported with a 95% confidence interval. The studies agreed that the ASES shoulder evaluation form is responsive and therefore useful instruments for detecting change in shoulder conditions.<sup>53-55</sup>

## **Summary**

Modern health care administration emphasizes the importance of the patient and their perspective on health and functioning as related to their specific injury or condition. Clinicians are able to measure a wide range of health status factors and monitor changes through the use of patient-reported outcome instruments. These instruments allow a clinician to assess the impact of a condition, evaluate treatment progress, and measure end results of services rendered from the perspective of the patient.<sup>2,10</sup>

Each PRO has an intended purpose and target population that it has been tested on and validated in. Careful consideration of the patient and their condition must be made by a clinician when selecting an appropriate instrument as the applicability of many PROs is limited to certain demographics. This may be of concern to clinicians who work with physically active individuals because many instruments have not been tested in these populations.

To date, little research has been done to test the applicability of PROs in physically active populations. Research into the relationship between scores on shoulder-specific outcome



measures and those on general health-related quality of life measures is also limited. The present study is specifically interested in the Short-Form Health Survey, the Disablement in the Physically Active Scale, the Disabilities of the Arm, Shoulder, and Hand, and the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form.

This study aimed to examine the effectiveness of patient-reported outcome measures in the physically active and to determine whether a relationship exists between general health-related quality of life and shoulder-specific outcome measures. The following research questions guided this study: (1) What is the relationship between scores on shoulder-specific and scores on general health-related quality of life patient-reported outcome measures? (2) Are patient-reported outcome measures appropriate in physically active populations?

## **CHAPTER III. METHODOLOGY**

### **Purpose**

The purpose of this study was to examine the effectiveness of patient-reported outcome measures in physically active populations and to determine whether a relationship exists between general health-related quality of life and shoulder-specific outcome measures. The following research questions guided this study: (1) What is the relationship between scores on shoulder-specific and scores on general health-related quality of life patient-reported outcome measures? (2) Are patient-reported outcome measures appropriate in physically active populations?

### **Research Design**

The design of this non-experimental study was survey based. Participants completed four different patient-reported outcome measures over the course of three sessions and the results were analyzed.

### **Sample of Study**

The participants sampled in this study were male and female athletes participating in various sports at North Dakota State University (NDSU) in Fargo, North Dakota. This group was selected because they are physically active and representative of varying levels and types of physical activity due to their participation in sports.

Permission was obtained from the Director of Athletics, the Director of Sports Medicine, and the Associate Director of Athletics – Internal Operations to conduct this study using NCAA Division I athletes as participants. Potential participants were recruited with the assistance of the athletic trainers of each sports team, and they were informed of the purpose and methodology of the study before involvement. Inclusion criteria for this study were: at least 18 years of age and current athlete on an NDSU sports team roster. Exclusion criteria were: current injury or history

of injury within the past month. The goal of this research study was to obtain data from at least 50 different participants.

### **Instrumentation**

The instruments used in this study were the Short-Form 36-Item Health Survey Version 1.0 (SF-36), the Disablement in the Physically Active Scale (DPA), the shortened Disabilities of the Arm, Shoulder, and Hand (QuickDASH), and the patient self-assessment section of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (p-ASES). Instruments can be found in Appendix A.

The SF-36 is a generic patient-reported outcome measure (PRO) designed to measure health-related quality of life (HRQOL) in a wide variety of populations. It is a 36-item questionnaire that covers eight dimensions of health: Physical Functioning, Social Functioning, Physical Role Limitations, Bodily Pain, General Mental Health, Emotional Role Limitations, Vitality, and General Health Perceptions.<sup>24,26</sup> Each dimension is scored and reported separately on a zero to 100 scale, where a higher score is representative of a more desirable health state.<sup>23,25</sup> Item responses are in the form of two-, three-, five-, and six-point Likert scales. Scoring of the SF-36 occurs in two steps. First, all responses are recoded based on the scoring key. Second, the statistical mean is taken for items in the same scale to create eight scale scores.

The DPA is a population-specific PRO designed to assess HRQOL in physically active individuals. Released in 2010, the DPA is a 16-item questionnaire that covers domains of impairment, functional limitation, disability, and quality of life.<sup>32,33</sup> Item responses are in the form of a 5-point Likert scale, where 1 indicates *no problem* and 5 indicates *severe problem*.<sup>33</sup> Scores are obtained by summing the responses and subtracting 16 so that a final score ranges from zero to 64.<sup>32</sup> Higher scores indicate increased disablement.

The QuickDASH is a shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire that is valid, reliable, and responsive and covers all of the same domains as the DASH.<sup>2,46-49</sup> The QuickDASH is a region-specific PRO for the upper extremity and contains 11 items which assess daily activity, symptoms, social function, work function, sleep, and confidence.<sup>34,36,38</sup> Items are scored on a 5-point Likert scale, where 1 indicates *no difficulty/symptoms* and 5 indicates *extreme difficulty/symptoms*.<sup>34,37,38</sup> A final score is obtained by taking the statistical mean of the responses, subtracting one, and multiplying by 25 to produce a score out of 100. Higher scores signify greater disability.

The ASES is a region-specific PRO used to assess shoulder function in two sections: a patient self-assessment section (p-ASES) and a physician evaluation section. The physician assessment is not used for scoring purposes so only the p-ASES was used in this study. The p-ASES is divided into three sections; there are six items related to pain, two items concerning shoulder instability, and 10 items related to activities of daily living (ADL) for each shoulder. The pain and instability sections contain yes/no questions and visual analog scales scored zero (best) to 10 (worst). The items in the ADL section are scored on a 4-point Likert scale where zero indicates *inability to perform* and three indicates *no difficulty*.<sup>50</sup> A final score is calculated using the formula:  $(10 - VAS_{\text{pain}}) * 5 + (5/3 * ADL_{\text{sum}})$ .<sup>50</sup> The p-ASES will produce a final score on a zero to 100 scale with higher scores being more desirable.

### **Procedure**

Institutional Review Board approval was obtained prior to data collection, which occurred over the course of a fall academic semester. Outcome measure scores were recorded at three time points: once preseason and twice during the season. Each time point was at least three weeks apart from the previous. An iPad was used for the administration of all outcome measures.

Prior to the first session, participants were informed of the purpose and nature of the study as well as where and when they would complete the outcome measures. Participants were read a script informing them of the purpose of the study, their rights and responsibilities for the study, and the procedure on how their responses would be collected. The script included a statement of consent which stated that by participating, all individuals consent the research team to use information provided in the study. The oral script can be found in Appendix B.

A demographic information questionnaire was administered prior to each data collection session. This questionnaire contained questions concerning age, sport participation, and injury history among others. A copy of this questionnaire is provided in Appendix A. The four outcome measures were administered in an order determined by Latin square. Participants were given as much time as necessary to fully complete each PRO. Data collection sessions lasted approximately 10 to 15 minutes for each participant. The first session occurred during preseason for the participants' respective sports. The second and third sessions occurred during the respective sports seasons with at least three weeks between each session.

A drawing was held after each of the three data collection sessions for everyone who participated in that session. Each drawing was for one of 5 ten dollar gift cards, and participants had three chances to win if they attended all three sessions.

### **Statistical Analysis**

Data processing was completed using IBM SPSS statistics software version 23. Descriptive statistics, including mean and standard deviation, were calculated for each of the four PROS for each testing session. A bivariate Pearson correlation was performed between scores on the four outcome measures with a significance level set at  $\alpha \leq .05$  to identify any relationships that might exist at each time-point. Additionally, a repeated-measures analysis of

variance (ANOVA) was calculated for each instrument to measure the effect of time on PRO scores. If a significant time effect was calculated, post hoc comparisons were performed using Bonferroni adjustment. Finally, test-retest reliability was determined using intraclass correlation coefficient (ICC) which was in turn used to calculate standard error of measurement (SEM) and minimal detectable change (MDC), the smallest amount of change necessary in a score to demonstrate true change beyond error.<sup>5,20</sup> SEM is calculated using the formula  $SEM = \text{pooled SD} * \sqrt{1-ICC}$  and MDC with a 95% confidence interval is calculated using the formula  $MDC = SEM * 1.96 * \sqrt{2}$ .

## CHAPTER IV. MANUSCRIPT

### Introduction

Evidence-based practice (EBP) is the foundation for health care administration in the sports medicine setting. EBP is the integration of a clinician's personal experience with the best available research and the values of their patients.<sup>1</sup> Clinical outcomes assessment enables EBP through the examination of patient values after medical interventions. It provides clinicians with means to assess treatment progress and measure the end results of services rendered which affect a patient's health-related quality of life (HRQOL).<sup>2</sup>

HRQOL is an individual's general feeling of overall well-being and encompasses physical, psychological, and social aspects of health as derived from personal beliefs, preferences, experiences, and expectations.<sup>3-8</sup> HRQOL is both individualistic and variable in its nature; it is influenced by injury, illness, and disease, as well as previous personal experiences and changing values and priorities.<sup>3</sup> Oftentimes, HRQOL is measured by using patient-reported outcome measures (PROs).

PROs are self-report questionnaires on a patient's health condition and include aspects such as symptoms, physical function, and general well-being.<sup>2,9</sup> PROs are either generic or specific in nature. Generic instruments measure overall wellbeing in a wide variety of patients, independent of injury or condition, whereas specific measures evaluate aspects related to a precise disease, injury, population, or anatomical region.<sup>2,3,8-10</sup> The target populations for outcome measures are typically as broad as possible, but often do not take into account that HRQOL varies between physically active and sedentary individuals. Many instruments are not valid in the physically active, or have not been tested in these populations, which may limit their use in a sports medicine setting.<sup>1,5,11-14</sup> Additionally, ceiling effects may be associated with the

use of some PROs on physically active individuals which may limit their validity in these populations. A ceiling effect is the level above which variance in an independent variable is no longer being measured and occurs when the best possible score is achieved by greater than 20% of participants at baseline.<sup>1,17</sup>

Studies have shown that quality of life differs in physically active individuals compared to the general population.<sup>11-15</sup> Many of these studies<sup>11,13,14</sup> assess HRQOL using the Short-Form 36-Item Health Survey (SF-36), a popular generic PRO. Fewer studies exist to analyze the results for physically active individuals on specific PRO instruments such as the Disablement in the Physically Active Scale (DPA), the shortened Disabilities of the Arm, Shoulder, and Hand questionnaire (QuickDASH), or the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES).

The purpose of this study was to examine the effectiveness of PRO measures in the physically active and to determine whether a relationship exists between general health-related quality of life and shoulder-specific outcome measures. The following research questions guided this study: (1) What is the relationship between scores on shoulder-specific and scores on general health-related quality of life patient-reported outcome measures? (2) Are patient-reported outcome measures appropriate in physically active populations?

## **Methodology**

### **Participants**

A total of 51 National Collegiate Athletic Association Division I athletes (38 women, 13 men) from multiple sports were initially included in this study. Participants' sports included cross country, football, soccer, softball, track and field, and volleyball. For the purposes of this study, sport participation qualified participants as being physically active. Participants were



excluded from this study if they were not at least 18 years of age or if they were not considered healthy at the first data collection session. Participants were considered healthy if they were cleared for sport activity without restrictions by their athletic trainer and did not self-report any concurrent injury or injury within the previous 30 days. Informed consent was obtained from all participants before the study. This study was approved by the university's institutional review board.

Six participants were disqualified from the study due to the exclusion criteria above and three participants dropped out of the study before completion. Therefore, data from 42 participants were available for analysis. Participants' demographic information can be found in Table 3.

**Table 3. Demographic Information**

<b>Sport</b>	<b>Sample</b>	<b>Age, y (Mean ± SD)</b>
M. Cross Country	3	19.33 ± 1.53
M. Football	2	20.00 ± 0.00
M. Track & Field	5	20.00 ± 2.00
<b>Men Total</b>	<b>10</b>	<b>19.80 ± 1.55</b>
W. Cross Country	5	18.80 ± 0.84
W. Soccer	8	19.88 ± 1.46
W. Softball	8	19.50 ± 0.93
W. Volleyball	8	18.88 ± 1.81
W. Track & Field	3	20.00 ± 1.00
<b>Women Total</b>	<b>32</b>	<b>19.38 ± 1.34</b>
<b>Overall Total</b>	<b>42</b>	<b>19.48 ± 1.38</b>

## Procedures

Participants were required to complete three data collection sessions over the course of a collegiate academic semester. The first time-point (T1) occurred during the preseason of the participants' respective sports while the second and third time-points (T2 and T3) occurred after respective sports seasons had started. At least three weeks were required to elapse between subsequent sessions to determine the effect of administration time point on PRO scores.

During each data-collection session, participants were required to fill out a demographic questionnaire and four PROs: the DPA, SF-36, QuickDASH, and p-ASES. Surveys were completed in-person on handheld tablets for the first session. Participants were provided an online link, to Qualtrics, to complete the surveys on their own for subsequent sessions.

## **Instrumentation**

### ***Disablement in the Physically Active Scale***

The DPA is a population-specific PRO designed to assess HRQOL in physically active individuals. Released in 2010, the DPA is a 16-item questionnaire that covers domains of impairment, functional limitation, disability, and quality of life.<sup>32,33</sup> Item responses are in the form of a 5-point Likert scale, where 1 indicates *no problem* and 5 indicates *severe problem*.<sup>33</sup> Scores range from zero to 100, where higher scores indicate increased disablement. The DPA has been shown to be valid and reliable (ICC = 0.943) in physically active patients with both acute ( $\alpha = .908$ ) and chronic ( $\alpha = .890$ ) injuries.<sup>32</sup>

### ***Short-Form 36-Item Health Survey***

The SF-36 is a generic PRO designed to measure HRQOL in a wide variety of populations. It is a 36-item questionnaire that covers eight dimensions of health: Physical Functioning, Social Functioning, Physical Role Limitations, Bodily Pain, General Mental Health, Emotional Role Limitations, Vitality, and General Health Perceptions.<sup>24,26</sup> Item responses are in the form of two-, three-, five-, and six-point Likert scales. Each dimension is scored and reported separately on a zero to 100 scale, where a higher score is representative of a more desirable health state.<sup>23,25</sup> The SF-36 has been demonstrated as valid and reliable with a range of ICC values between 0.63 (emotional role limitations) and 0.81 (physical functioning).<sup>26</sup>

### ***Shortened Disabilities of the Arm, Shoulder, and Hand***

The QuickDASH is a shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire. It contains 11 items, which assess daily activity, symptoms, social function, work function, sleep, and confidence.<sup>34,36,38</sup> Items are scored on a 5-point Likert scale, where 1 indicates *no difficulty/ symptoms* and 5 indicates *extreme difficulty/ symptoms*.<sup>34,37,38</sup> Scores are reported on a scale from zero to 100, where higher scores signify greater disability. The QuickDASH has been shown to be valid and reliable (ICC = 0.93) for use on patients with injuries to the upper extremity.<sup>2,46-49</sup>

### ***American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form***

The ASES is a region-specific PRO used to assess shoulder function in two sections: a patient self-assessment section (p-ASES) and a physician evaluation section.<sup>53</sup> The physician assessment is not used for scoring purposes, so it was not used in this study. The p-ASES is divided into three sections; there are six items related to pain, two items concerning shoulder instability, and 10 items related to activities of daily living (ADL) for each shoulder. The pain and instability sections contain yes/ no questions and visual analog scales scored zero (best) to 10 (worst). The items in the ADL section are scored on a 4-point Likert scale where zero indicates *inability to perform* and three indicates *no difficulty*.<sup>50</sup> The p-ASES will produce a final score on a zero to 100 scale with higher scores being more desirable. The ASES has been shown to be a valid and reliable (ICC = 0.84) outcome measure.<sup>53</sup>

### **Statistical Analysis**

Data processing was completed using IBM SPSS statistics software version 23. Descriptive statistics, including mean and standard deviation, were calculated for each of the four PROs for each testing session. A bivariate Pearson correlation was performed between

scores on the four outcome measures with a significance level set at  $\alpha \leq .05$  to identify any relationships that might exist at each time-point. Additionally, a repeated-measures analysis of variance (ANOVA) was calculated for each instrument to measure the effect of time on PRO scores. If a significant time effect was calculated, post hoc comparisons were performed using Bonferroni adjustment. Finally, test-retest reliability was determined using intraclass correlation coefficient (ICC) which was in turn used to calculate standard error of measurement (SEM) and minimal detectable change (MDC), the smallest amount of change necessary in a score to demonstrate true change beyond error.<sup>5,20</sup> SEM is calculated using the formula  $SEM = \text{pooled SD} * \sqrt{(1-ICC)}$  and MDC with a 95% confidence interval is calculated using the formula  $MDC = SEM * 1.96 * \sqrt{2}$ . ICC values were interpreted as either weak ( $\leq .20$ ), moderate ( $.20 - .74$ ), or strong ( $\geq .75$ ).<sup>56</sup>

## Results

**Table 4. Patient-Reported Outcome Measure Scores (n = 42)**

Instrument	Subscale	Mean $\pm$ SD		
		T1	T2	T3
<b>DPA</b>		9.33 $\pm$ 8.30	8.86 $\pm$ 8.19	6.26 $\pm$ 7.68
<b>SF-36</b>	Vitality	70.24 $\pm$ 10.87	66.79 $\pm$ 13.43	69.52 $\pm$ 12.92
	PF	98.21 $\pm$ 4.11	97.38 $\pm$ 7.75	99.17 $\pm$ 3.97
	Pain	78.01 $\pm$ 13.25	81.07 $\pm$ 14.60	83.15 $\pm$ 14.21
	GHP	75.50 $\pm$ 14.82	74.41 $\pm$ 11.95	75.40 $\pm$ 13.46
	RLP	95.24 $\pm$ 9.94	95.83 $\pm$ 12.24	98.81 $\pm$ 7.72
	RLE	93.65 $\pm$ 19.81	95.24 $\pm$ 13.91	92.86 $\pm$ 21.51
	SF	90.77 $\pm$ 16.11	94.05 $\pm$ 11.79	93.75 $\pm$ 11.47
	GMH	82.29 $\pm$ 9.56	82.76 $\pm$ 10.43	80.57 $\pm$ 13.18
<b>QuickDASH</b>		3.03 $\pm$ 4.81	1.95 $\pm$ 3.16	0.97 $\pm$ 2.41
<b>ASES</b>		92.84 $\pm$ 8.02	95.66 $\pm$ 6.52	96.92 $\pm$ 5.01

### Relationships between PRO Scores

The original sample of 42 participants (10 men, 32 women) was included in this analysis. Descriptive statistics for each PRO can be found in Table 4. Pearson correlations were calculated

for each of the relationships between mean PRO scores at each time point. These values can be found in Table 5. In total, 20 significant relationships were calculated across the three time points. Significant relationships that were consistent across all three of the time points were between the DPA and the SF-36 Physical Function (T1  $r = -.465$ ,  $P = .002$ ; T2  $r = -.531$ ,  $P = .000$ ; T3  $r = -.533$ ,  $P = .000$ ) and Pain (T1  $r = -.543$ ,  $P = .000$ ; T2  $r = -.704$ ,  $P = .000$ ; T3  $r = -.670$ ,  $P = .000$ ) subscales. Significant relationships consistent across two of the three time points existed between the DPA and the SF-36 General Health Perceptions (T2  $r = -.368$ ,  $P = .017$ ; T3  $r = -.375$ ,  $P = .014$ ) subscale as well as between the QuickDASH and the ASES (T2  $r = -.335$ ,  $P = .030$ ; T3  $r = -.434$ ,  $P = .002$ ).

**Table 5. Relationships between Patient-Reported Outcome Measure Scores (n = 42)**

Instrument	Subscale	Pearson Correlation (r)								
		T1			T2			T3		
		DPA	QDASH	ASES	DPA	QDASH	ASES	DPA	QDASH	ASES
SF-36	Vitality	-.194	.209	.054	-.105	-.208	.117	-.013	-.145	.013
	PF	-.465 <sup>a</sup>	.000	-.062	-.531 <sup>a</sup>	-.183	-.219	-.533 <sup>a</sup>	-.347 <sup>b</sup>	.277
	Pain	-.543 <sup>a</sup>	-.070	.109	-.704 <sup>a</sup>	-.278	.047	-.670 <sup>a</sup>	-.306 <sup>b</sup>	.307 <sup>b</sup>
	GHP	-.284	-.005	.126	-.368 <sup>b</sup>	-.305 <sup>b</sup>	.460 <sup>a</sup>	-.375 <sup>b</sup>	-.176	-.097
	RLP	-.143	-.097	-.011	-.450 <sup>a</sup>	-.394 <sup>a</sup>	.073	-.221	.064	-.097
	RLE	-.185	-.084	-.080	-.399 <sup>a</sup>	-.246	.013	-.072	.030	-.159
	SF	-.068	-.024	-.048	-.236	-.221	.438 <sup>a</sup>	-.095	-.025	-.219
	GMH	-.272	-.087	.068	-.209	-.282	.339 <sup>b</sup>	.076	.031	-.023
QDASH		-.044		.172			.246			
ASES		-.059	-.233		-.002	-.335 <sup>b</sup>		.014	-.464 <sup>a</sup>	

<sup>a</sup> Significant relationship ( $P \leq .01$ ).

<sup>b</sup> Significant relationship ( $P \leq .05$ ).

## Appropriateness of PROs in Physically Active Populations

### *Effect of Administration Time Point on PRO Scores*

Seven participants reported sustaining an injury after data collection had begun, so their data was not included in this analysis. A total of 35 participants (9 men, 26 women) were therefore included in this analysis. Descriptive statistics for these participants can be found in Table 6. A time effect was calculated for the DPA ( $P = .001$ ) with scores significantly lower at T3 than at T1 ( $P = .002$ ). A time effect was observed for the SF-36 Pain subscale ( $P = .049$ ), but

post hoc analyses were not significant. Time effects were not observed for any of the other scores.

***Test-Retest Reliability and Minimal Detectable Change***

A total of 35 participants (9 men, 26 women) were included in this analysis. Reliability coefficients and MDCs for each PRO can be found in Table 7.

**Table 6. PRO Scores and Repeated-Measures ANOVA (n = 35)**

Instrument	Subscale	Mean ± SD			Sig.
		T1	T2	T3	
<b>DPA</b>		9.03 ± 8.14	6.29 ± 5.72	4.94 ± 6.21	.001 <sup>a</sup>
<b>SF-36</b>	Vitality	70.29 ± 11.69	66.29 ± 13.41	68.86 ± 12.84	.209
	PF	98.71 ± 3.05	98.57 ± 4.94	99.14 ± 4.29	.472
	Pain	79.21 ± 12.64	83.50 ± 10.66	84.36 ± 12.94	.049 <sup>b</sup>
	GHP	76.43 ± 15.85	74.64 ± 12.76	75.36 ± 13.94	.686
	RLP	95.71 ± 9.56	97.14 ± 10.09	100.00 ± 0.00	.088
	RLE	92.38 ± 21.52	97.14 ± 9.47	91.43 ± 23.35	.382
	SF	91.07 ± 15.63	93.93 ± 12.27	93.57 ± 11.89	.497
	GMH	82.06 ± 10.28	82.63 ± 10.67	79.66 ± 13.80	.339
<b>QuickDASH</b>		3.12 ± 4.90	2.08 ± 3.28	1.10 ± 2.61	.075
<b>ASES</b>		93.10 ± 7.47	94.81 ± 6.85	96.60 ± 5.21	.055

<sup>a</sup> T1 different from T3 (P = .002)

<sup>b</sup> Bonferroni correction insignificant

**Table 7. Test-Retest Reliability and Minimal Detectable Change (n = 35)**

Instrument	Subscale	ICC	SEM	95% MDC
<b>DPA</b>		.603	4.27	11.83
<b>SF-36</b>	Vitality	.441	9.47	26.25
	PF	.765	2.02	5.60
	Pain	.427	9.18	25.44
	GHP	.573	9.30	25.79
	RLP	.060	7.78	21.57
	RLE	.044	18.71	51.85
	SF	.328	10.96	30.37
	GMH	.386	9.16	25.39
<b>QuickDASH</b>		.100	3.53	9.79
<b>ASES</b>		.182	5.95	16.49

## **Discussion**

The purpose of this study was to examine relationships between scores on general and shoulder-specific PROs and to determine whether these measures are appropriate for use in physically active populations. Significant relationships were found between the DPA and several subscales of the SF-36 across multiple time points, and significant time effects were measured for the DPA and the SF-36 Pain subscale. Test-retest reliability was adequate for the DPA and some SF-36 subscales.

### **Relationships between PRO Scores**

Moderate relationships were demonstrated between the DPA and the SF-36 Physical Function subscale while moderate to strong relationships were demonstrated between the DPA and the SF-36 Pain subscale across all three time points. Weak relationships were found between the DPA and the SF-36 General Health Perceptions subscale across only two of the three time points. Additionally, weak to moderate relationships were found between the QuickDASH and the ASES across two of the three time points. Other relationships between PRO scores existed but they were not consistent across the multiple time points. These results suggest that there may be some overlap in the DPA domains and the SF-36 subscales; however, the SF-36 is more suitable for measuring aspects related to mental health and emotional well-being.

When comparing the generic DPA and SF-36 outcome measures to the shoulder region-specific QuickDASH and ASES outcome measures, few weak to moderate statistically significant relationships existed. However, these relationships only existed at one of the three time points. This lack of consistency, as well as the remaining insignificant relationships identified, suggests that clinicians should use both generic and region-specific instruments to

assess HRQOL in their patients. A recent study<sup>5</sup> demonstrated similar findings when examining relationships between the DPA and two lower-extremity PROs in a population of female soccer players. It is unclear what kind of effect upper extremity injuries would have had on the relationships between these instruments in these populations.

Significant ceiling effects were found in the SF-36 physical function, physical role limitations, emotional role limitations, and social function subscales as well as in the QuickDASH and the ASES. Ceiling effects may have limited the validity of the SF-36 Physical Function (76.2% achieved perfect score at T1), Physical Role Limitations (81.0%), Emotional Role Limitations (88.1%), and Social Function (66.7%) subscales as well as the QuickDASH (52.4%) and the ASES (42.9%) when used on this sample of physically active individuals. These results suggest that clinicians should consider the DPA for measuring impairment, function, and disability in their physically active patients, but may want to consider the SF-36 for measuring aspects of mental health until a suitable alternative is developed for physically active populations. For injuries to the upper extremity, clinicians should continue to use both generic and region-specific PROs to accurately measure HRQOL in their patients.

### **Appropriateness of PROs in Physically Active Populations**

#### ***Effect of Administration Time Point on PRO Scores***

A main effect of time was demonstrated for the DPA in this study, with a difference between the T1 and T2 time points. Additionally, a main effect of time was present for the SF-36 Pain subscale, though post hoc analysis did not reveal any differences between time points. The difference between time points that was observed for the DPA did not exceed its MDC that was found in this study, which may indicate that the difference was due to error of the measure rather than a true change occurring. Based on these results, administration time point did not have an



effect on scores for any of the PRO measures. Changes in scores over time are therefore unlikely to be related to an individual's participation in physical activity; aspects of health measured by the DPA, the SF-36, the QuickDASH, and the ASES were not impacted by participants' physical activity in this study. Clinicians can expect PRO scores to remain constant throughout an athletic season provided outside factors, such as injury, do not occur. Additionally, baseline scores, which are typically measured pre-season, can be obtained in-season if the clinician deems it necessary.

### ***Test-Retest Reliability and MDC***

Moderate test-retest reliability was found for the DPA as well as for the SF-36 Vitality, Pain, General Health Perceptions, Social Function, and General Mental Health subscales. Strong test-retest reliability was found for the SF-36 Physical Function subscale. The remaining SF-36 subscales, the QuickDASH, and the ASES were interpreted as having weak test-retest reliability. From a reliability standpoint, the DPA and the SF-36 Vitality, Pain, General Health Perceptions, Social Function, and General Mental Health subscales were adequate for use in physically active populations. Previously, an ICC of .943 was reported<sup>32</sup> on the DPA in a sample of 386 competitive and recreational athletes. A later study<sup>5</sup> reported an ICC of .792 on the DPA in a sample of 16 female collegiate soccer players, with a 12.48 MDC. In a study on the SF-36<sup>26</sup>, ICC values for the SF-36 were found to be between .60 (Social Function) and .81 (Physical Function). Test-retest reliability was previously reported as excellent for both the QuickDASH<sup>47</sup> and the ASES<sup>53</sup> on patients with shoulder dysfunctions.

Some factors were identified which may have led to the discrepancies in this data compared to those found previously published studies. An average of  $35.05 \pm 12.05$  days elapsed between data collection sessions, which is much greater than time periods used in previous

studies. Time frames found in other studies of test-retest reliability of these PROs ranged between one and 25 days on average.<sup>5,47,53</sup> The increased time between data collection in this study allowed for a greater possibility for the introduction of uncontrolled variables. Another factor, which may limit reliability in this study, is that the SF-36, the QuickDASH, and the ASES were not designed with the high level of physical function commonly found with athletes in mind. While some of the SF-36 subscales were found to have moderate to strong test-retest reliability, its overall reliability when used on physically active individuals may be limited by the other subscales.

### **Limitations**

This study had its limitations. First, all participants were competitive athletes at an NCAA Division I university. It is unclear whether these results can be generalized to recreational or adolescent athletes. Additionally, this study investigated two shoulder-specific PROs and their applicability in physically active individuals regardless of the type of activity. The QuickDASH and the ASES were designed for use on patients with shoulder and other upper extremity dysfunctions, while this study used mostly healthy participants or those with non-shoulder injuries. It is also possible that some of the items from these instruments did not apply to participants who do not use the upper extremity in their sport, such as soccer or cross country. Further investigation may be warranted to determine the effects of these instruments in exclusively overhead athletes. Future research could also focus on the effects that upper extremity injuries have on scores to these instruments.

### **Conclusions**

The DPA and several subscales of the SF-36 may be effective measures of HRQOL in physically active populations, though significant ceiling effects may limit the validity of the SF-

36 in these populations. In addition, participation in physical activity across several months of various sports seasons did not affect scores on the DPA, the SF-36, the QuickDASH, or the ASES. Changes in scores on these measures are therefore likely to come from other sources, such as injury. Further research may be necessary before any conclusions can be made regarding the efficacy of the QuickDASH and the ASES in the physically active. Clinicians are urged to consider using both generic and region-specific PROs on their patients to sufficiently measure all aspects of HRQOL.

## REFERENCES

1. Denegar CR, Vela LI, Evans TA. Evidence-based sports medicine: Outcomes instruments for active populations. *Clinics in Sports Medicine*. July 2008; 27(3): 339-351.
2. Valovich McLeod TC, Snyder AR, Parsons JT, Curtis Bay R, Michener LA, Sauers EL. Using disablement models and clinical outcomes assessment to enable evidence-based athletic training practice, part II: Clinical outcomes assessment. *Journal of Athletic Training*. July 2008; 43(4): 437-445.
3. Parsons JT, Snyder AR. Health-related quality of life as a primary clinical outcome in sport rehabilitation. *Journal of Sports Rehabilitation*. February 2011; 20(1): 17-36.
4. Cieza A, Stucki G. Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Quality of Life Research*. June 2005; 14(5): 1225-1237.
5. Hoch JM, Druvenga B, Ferguson BA, Houston MN, Hoch MC. Patient-reported outcomes in male and female collegiate soccer players during an athletic season. *Journal of Athletic Training*. September 2015; 50(9): 930-936.
6. Houston MN, Hoch MC, Hoch JM. Health-related quality of life in athletes: A systematic review with meta-analysis. *Journal of Athletic Training*. June 2016; 51(6): 442-453.
7. Houston MN, Hoch JM, Van Lunen BL, Hoch MC. The development of summary components for the Disablement in the Physically Active scale in collegiate athletes. *Quality of Life Research*. November 2015; 24(11): 2657-2662.
8. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *The New England Journal of Medicine*. March 1996; 334(13): 835-840.
9. Kyte DG, Calvert M, van der Wees PJ, ten Hove R, Tolan S, Hill JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*. June 2015; 101(2): 119-125.
10. Beaton DE, Schemitsch E. Measures of health-related quality of life and physical function. *Clinical Orthopaedics & Related Research*. August 2003; 413: 90-105.
11. Huffman GR, Park J, Roser-Jones C, Sennett BJ, Yagnik G, Webner D. Normative SF-36 values in competing NCAA intercollegiate athletes differ from values in the general population. *The Journal of Bone and Joint Surgery*. March 2008; 90(3): 471-476.
12. Lam KC, Snyder Valier AR, Bay RC, Valovich McLeod TC. A unique patient population? Health-related quality of life in adolescent athletes versus general, healthy adolescent individuals. *Journal of Athletic Training*. March 2013; 48(2): 233-241.
13. McAllister DR, Motamedi AR, Hame SL, Shapiro MS, Dorey FJ. Quality of life assessment in elite collegiate athletes. *American Journal of Sports Medicine*. November 2001; 29(6): 806-810.
14. Snyder AR, Martinez JC, Bay RC, Parsons JT, Sauers EL, Valovich McLeod TC. Health-related quality of life differs between adolescent athletes and adolescent nonathletes. *Journal of Sport Rehabilitation*. August 2010; 19(3): 237-248.
15. Brown DW, Brown DR, Heath GW, Balluz L, Giles WH, Ford ES, Mokdad AH. Associations between physical activity dose and health-related quality of life. *Medicine & Science in Sports & Exercise*. May 2004; 36(5): 890-896.
16. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health*. November 2007; 10(s2): s94-s105.

17. Hsu JE, Nacke E, Park MJ, Sennett BJ, Huffman GR. The Disabilities of the Arm, Shoulder, and Hand questionnaire in intercollegiate athletes: Validity limited by ceiling effect. *Journal of Shoulder and Elbow Surgery*. April 2010; 19(3): 349-354.
18. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*. October 1998; 2(14): 19-43.
19. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. *Journal of Psychomatic Research*. April 2010; 68(4): 319-323.
20. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*. February 2008; 61(2): 102-109.
21. Turner D, Shünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, Guyatt GH. The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology*. January 2010; 63(1): 28-36.
22. Duncan EA, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: A systematic review. *BMC Health Services Research*. May 2012; 12(96).
23. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopedics: A brief guide. *Journal of Bone & Joint Surgery*. October 2015; 97(19): 1628-1634.
24. Ware JE, Shelbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Medical Care*. June 1992; 30(6): 473-483.
25. Ware JE. SF-36 Health Survey update. *Spine*. December 2000; 25(24): 3130-3139.
26. Brazier JE, Harper R, Jones NMB, O’Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. July 1992; 305: 160-164.
27. Hawthorne G, Osborne RH, Taylor A, Sansoni J. The SF36 Version 2: Critical analyses of population weights, scoring algorithms and population norms. *Quality of Life Research*. May 2007; 16(4): 661-673.
28. Jenkinson C, Coulter A, Wright L. Short form 36 (SF 36) health survey questionnaire: normative data for adults of working age. *BMJ*. May 1993; 306: 1437-1440.
29. Jenkinson C, Wright L, Coulter A. Criterion validity and reliability of the SF-36 in a population sample. *Quality of Life Research*. February 1994; 3(1): 7-12.
30. McHorney CA, Ware Jr JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*. March 1993; 31(3): 247-263.
31. McHorney CA, Ware Jr JE, Raczek Lu JF, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*. January 1994; 32(1): 40-66.
32. Vela LI, Denegar CR. The Disablement in the Physically Active Scale, part II: The psychometric properties of an outcomes scale for musculoskeletal injuries. *Journal of Athletic Training*. November 2010; 45(6): 630-641.
33. Vela LI, Denegar C. Transient disablement in the physically active with musculoskeletal injuries, part I: A descriptive model. *Journal of Athletic Training*. November 2010; 45(6): 615-629.
34. Beaton DE, Davis AM, Hudak P, McConnell S. The DASH (Disabilities of the Arm, Shoulder and Hand) Outcome Measure: What do we know about it now? *Hand Therapy*. December 2001; 6(4): 109-118.

35. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *Journal of Hand Therapy*. April 2001; 14(2): 128-146.
36. Bialocerkowski A. Disabilities of the Arms, Shoulder and Hand questionnaire. *Australian Journal of Physiotherapy*. 2007; 53(2): 135.
37. Franchignoni F, Giordano A, Sartorio F, Vercelli S, Pascariello B, Ferriero G. Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): A factor analysis and Rasch validation study. *Archives of Physical Medicine & Rehabilitation*. September 2010; 91(9): 1370-1377.
38. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: The DASH (Disabilities of the Arm, Shoulder, and Hand). *American Journal of Industrial Medicine*. June 1996; 29(6): 602-608.
39. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function. *Arthritis Care & Research*. November 2011; 63(S11): S174-S188.
40. Clarke MG, Dewing CB, Schroder DT, Solomon DJ, Provencher MT. Normal shoulder outcome score values in the young, active adult. *Journal of Shoulder and Elbow Surgery*. May 2009; 18(3): 424-428.
41. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American Academy of Orthopaedic Surgeons outcomes instruments: normative values from the general population. *The Journal of Bone and Joint Surgery*. February 2002; 84(2): 208-215.
42. Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskeletal Disorders*. June 2003; 4(11).
43. Michener LA, Leggin BG. A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *Journal of Hand Therapy*. April 2001; 14(2): 68-76.
44. Slobogean GP, Noonan VK, O'Brien PJ. The reliability and validity of the Disabilities of the Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and Short Form-6D outcome instruments in patients with proximal humeral fractures. *Journal of Shoulder and Elbow Surgery*. April 2010; 19(3): 342-348.
45. Beaton DE, Wright JG, Katz JN, Amadio P, Bombardier C, Cole D, Davis A, Hudak P, Marx R, Hawker G, Makela M, Punnett L. Development of the QuickDASH: Comparison of three item-reduction approaches. *The Journal of Bone & Joint Surgery*. May 2005; 87(5): 1038-1046.
46. Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (Quick DASH): Validity and reliability based on responses within the full-length DASH. *BMC Musculoskeletal Disorders*. May 2006; 7(44).
47. 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. *Journal of Shoulder and Elbow Surgery*. November 2009; 18(6): 920-926.
48. Franchignoni F, Ferriero G, Giordano A, Sartorio F, Vercelli S, Brigatti E. Psychometric properties of QuickDASH – A classical test theory and Rasch analysis study. *Manual Therapy*. April 2011; 16(2): 177-182.

49. Polson K, Reid D, McNair PJ, Larmer P. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. *Manual Therapy*. August 2010; 15(4): 404-407.
50. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, Iannotti JP, Mow VC, Sidles JA, Zuckerman JD. A standardized method for the assessment of shoulder function. *Journal of Shoulder and Elbow Surgery*. November 1994; 3(6): 347-352.
51. Brinker MR, Cuomo JS, Popham GJ, O'Connor DP, Barrack RL. An examination of bias in shoulder scoring instruments among healthy collegiate and recreational athletes. *Journal of Shoulder and Elbow Surgery*. September 2002; 11(5): 463-469.
52. Sallay PI, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. *Journal of Shoulder and Elbow Surgery*. November 2003; 12(6): 622-627.
53. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: Reliability, validity, and responsiveness. *Journal of Shoulder and Elbow Surgery*. November 2002; 11(6): 587-594.
54. Oh JH, Jo KH, Kim WS, Gong HS, Han SG, Kim YH. Comparative evaluation of the measurement properties of the various shoulder outcome instruments. *The American Journal of Sports Medicine*. June 2009; 37(6): 1161-1168.
55. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *Journal of Shoulder and Elbow Surgery*. November 1998; 7(6): 565-572.
56. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3<sup>rd</sup> ed. Upper Saddle River, NJ: Prentice Hall; 2009.

## APPENDIX A. DEMOGRAPHIC QUESTIONNAIRE

**NDSU**

NORTH DAKOTA STATE UNIVERSITY

### Demographic Questionnaire

This survey is intended to gather some background information on participants of the study. Please answer the following questions as accurately as possible. If there is any information you do not wish to disclose, please leave the question blank.

1. Date of Birth (MM/DD/YYYY)

\_\_\_\_\_

2. Year in School (circle one)

1    2    3    4    5    6+

3. Biological Sex (circle one)

Male          Female

4. Is English your primary language? (circle one)

Yes          No

5. What sport do you play at NDSU?

\_\_\_\_\_

6. How many years have you played your sport?

1    2    3    4    5    6    7    8    Other: \_\_\_\_\_

7. Are you currently experiencing any shoulder pain?

Yes          No

If yes, which shoulder?

Right          Left          Both

8. Do you currently have any injury (to any body part) that prevents you from playing your sport or limits your participation in your sport in some way?

Yes          No

If yes, please describe your injury

\_\_\_\_\_

9. Have you experienced any injury within the past 30 days that no longer bothers you?

Yes          No

10. Have you had any upper extremity surgery in your lifetime?

Yes          No

If yes, please describe

\_\_\_\_\_

11. What is your dominant hand? (For baseball and softball players, if you bat and throw with different hands, please choose *ambidextrous*)

Right          Left          Ambidextrous



## **APPENDIX B. ORAL SCRIPT FOR RESEARCHER**

### **[Researcher Reads]**

Good afternoon,

My name is Chris Peroutka; I am a Certified Athletic Trainer and graduate student at North Dakota State University. As part of my master's thesis I am conducting research on patient-reported outcome measures and their use in physically active populations.

The purpose of this study is to examine the effectiveness of patient-reported outcome measures in a physically active population and to determine whether a relationship exists between general health-related quality of life and shoulder-specific outcome measures. Patient-reported outcome measures are questionnaires on a patient's health condition which often includes aspects such as symptoms, physical function, and general well-being. Many measures have not been fully evaluated in physically active populations which may limit their use in the athletic training clinical setting. Your participation is requested to be representative of a varying levels and types of physical activity due to your participation in competitive sports.

This study will consist of five questionnaires in total: one demographic questionnaire and four different patient-reported outcome measures. The outcome measures that will be used are the Short-Form 36-Item Health Survey, the Disablement in the Physically Active Scale, the Disabilities of the Arm, Shoulder, and Hand questionnaire, and the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form. Your responses are entirely confidential. This study is being conducted entirely by an outside entity so there will be no retribution for your participation or responses. None of the information provided will be used to identify you and no one other than myself will have access to your responses. Also, responses will not be individually reported, but analyzed and reported as a whole.

Are there any questions at this time?

By completing these questionnaires, you agree to participate in the study and grant consent for the research team to use information provided in the questionnaires, to include for publication purposes. There will be three total sessions in which I ask you to complete the questionnaires, with each session at least three weeks after the previous.

A drawing will be held after each of the three data collection sessions for everyone who participated in that session. Each drawing will be for one of ten \$5 gift cards, and you will have three chances to win one (or multiple) if you attend all three sessions. You will have an approximately 10-20% chance to win a gift card each session (depending on total participants for each session).

If you are not interested in participating you are free to leave at this time.

**[Those not interested will leave now]**

For your part, I ask that you read the directions thoroughly before beginning each questionnaire. Please provide the most accurate responses possible and try not to leave any question blank. Your participation is entirely voluntary. You can quit at any time once questionnaire administration has begun.

If you have any questions or concerns about research participants rights or to file a complaint regarding the research you can reach the NDSU Human Research Protection Office at (701) 231-8908 or toll-free at 1-855-800-6717.

Additionally, if you have any questions or concerns about your participation in this study I can be contacted at the following sources; christopher.peroutka@ndsu.edu, (608)445-3413.

Are there any questions or concerns about our expectations for this study?

When I say, you may begin. If there is a technological issue with your iPad, please let me know so I can fix it before you proceed. When you are finished, please bring your iPad to me and you will be free to go.

If there are no other questions at this time, you may begin.

**[After participant finishes questionnaires, we will discuss when they are able to come in for the next data collection session, at least three weeks later]**