Patient-Derived Outcomes Assessment: Psychometric Validation of Outcomes Instruments, Patient Satisfaction with Outcome, and Expected Value Decision Analysis

Mininder S. Kocher, MD, MPH

Department of Orthopaedic Surgery, Children's Hospital

ABSTRACT

Background: The emphasis of the outcomes assessment movement has been patient-derived outcomes assessment. The goal of this body of work was to advance patient-derived outcomes assessment in orthopaedic surgery. Specifically, the purposes of these studies were: (Part 1) to psychometrically validate commonly used condition-specific outcome instruments, (Part 2) to determine the predictors of patient satisfaction with outcome after surgery, and (Part 3) to utilize patient preferences in understanding decision-making under conditions of uncertainty.

Methods: (Part 1): Test-retest reliability, internal consistency, content-criterion-construct validity, and responsiveness to change were determined for the Lysholm Knee Score for chondral disorders and meniscal disorders, and for the ASES Shoulder score for instability, rotator cuff disease, and glenohumeral arthritis. (Part 2): The predictors of patient satisfaction with outcome after ACL reconstruction and rotator cuff surgery were determined. (Part 3): Expected-value decision analysis was used to explore decision-making for management of Achilles tendon rupture and the contralateral hip in SCFE.

Results: (Part 1): The Lysholm Knee Score and the ASES Shoulder Score demonstrate adequate pyschometric performance. (Part 2): Specific subjective and objective parameters were identified as predictors of patient satisfaction with outcome after ACL and rotator cuff surgery. (Part 3): Surgical management of Achilles tendon rupture and observation of the contralateral hip in SCFE are the preferred treatment strategies.

Discussion: Validation of outcomes intruments, identifying determinants of patient-satisfaction, and exploring patient preferrences in decision-making are important methodologies for advancement of patient-derived outcomes assessment in orthopaedics.

Address Correspondence to:

Mininder S. Kocher, MD, MPH
Department of Orthopaedic Surgery
Children's Hospital
Harvard Medical School
Harvard School of Public Health
300 Longwood Avenue
Boston, MA 02115

Phone: (617) 355-7497
Fax: (617) 730-0465
Email: **mininder.kocher@childrens.harvard.edu**

INTRODUCTION

Assessment and accountability have been called the "third revolution" in the medical care of patients (1). The health care outcomes assessment movement has been propelled by contemporaneous societal, technological, and economical forces including increased oversight, large small-area variations in the delivery of care, lack of evidence-based practice, and escalating costs of health care (2). Outcomes research has thus become the pivotal means to assessment and accountability. The emphasis of the outcomes assessment movement has been patient-derived outcomes assessment. This emphasis has been a paridigm shift in assessment from physician-derived assessment.

This manuscript overviews a body of work involving patient-derived outcomes assessment in orthopaedic surgery. In Part 1, the psychometric properties of reliability, validity, and responsiveness were established for two frequently used condition-specific outcomes instruments, the Lysholm Knee Score and the ASES Shoulder Score (3,4,5). In Part 2, predictors of patient satisfaction with outcome after ACL reconstruction and after rotator cuff surgery were determined (6,7). In Part 3, expected-value decision analysis was used to explore decision-making for management of Achilles tendon rupture and management of the contralateral hip in SCFE (8,9).

Part 1:

The use of outcome instruments whose psychometric properties have been vigorously established is essential. The majority of condition-specific outcomes instruments in orthopaedic surgery were created empirically and have not been been psychometrically validated. The important psychometric properties of an outcome instrument include reliability, validity, and responsiveness. Reliability refers to the reproducibility of an outcome measure, either between subjects (test-retest reliability) or between observers (interobserver reliability). Validity questions whether an outcome instrument actually measures what it intends to measure. Components of validity include content validity ("face" validity and floor/ceiling effects), criterion validity (how an instrument compares to an accepted "gold" standard instrument), and construct validity (does the instrument follow expected noncontroversial hypotheses). Responsiveness assesses changes in the instrument value over time or treatment.

The management of chondral injuries in the knee has received much attention with techniques such as abrasion arthroplasty, microfracture, autologous osteochondral grafting, allogeneic osteochondral grafting, periosteal grafting, and autologous chondrocyte implantation. Outcomes assessment after the treatment of chondral disorders of the knee has been very variable and has involved the use of various outcome instruments, such as the International Knee Documentation Committee (IKDC) form, the Tegner activity scale, the Cincinnati knee scale, the Hospital for Special Surgery knee scale, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), the Knee Society knee scale, and the Lysholm knee scale. Recently, the International Cartilage Repair Society folded its assessment documentation into the new IKDC form.

Injuries to the meniscus cartilage of the knee are common. A torn meniscus is the most common reason to undergo arthroscopic surgery. Outcomes assessment after the treatment of meniscus injuries of the knee has involved the use of various outcome instruments, such as the Tapper and Hoover system, Knee Injury and Osteoarthritis Outcome Score (KOOS), Lysholm knee score, International Knee Documentation Committee(IKDC) score, the Cincinnati Knee Rating Scale, and Tegner activity scale.

The Lysholm knee scale is a condition-specific outcome measure that contains eight domains: limp, locking, pain, stair climbing, support, instability, swelling, and squatting. An overall score of 0-100 is calculated, and is graded: excellent: 95-100, good: 84-94, fair: 65-83, and poor <65. Originally designed for assessment of ligament injuries of the knee, the Lysholm knee scale has been utilized for a variety of knee conditions, including chondral disorders of the knee.

Outcomes assessment after the treatment of shoulder disorders has typically involved the use of various condition-specific or region-specific outcome instruments, such as the Constant-Murley Shoulder Score, the University of California at Los Angeles (UCLA) Shoulder Score, the Rowe Score, the Simple Shoulder Test (SST), the Shoulder Pain and Disability Index (SPADI), the Disabilities of the Arm, Shoulder, and Hand Module (DASH), the Oxford Shoulder Score (OSS), the L'Insalata Shoulder Rating Questionnaire (LSRQ), the University of Pennsylvania Shoulder Score, the Shoulder Severity Index, the Western Ontario Rotator Cuff Index, the Western Ontario Osteoarthritis of the Shoulder (WOOS) index, the Western Ontario Shoulder Instability Index (WOSI), and the Rotator Cuff Quality of Life (RC-QOL) measure.

The American Shoulder and Elbow Surgeons (ASES) shoulder scale was published in 1994 by the Research Committee of the American Shoulder and Elbow Surgeons and contains both a patient-derived subjective assessment and a physician-derived objective assessment. The subjective patient self-report section consists of two equally weighted domains, pain and function, and has been widely used for outcomes assessment in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. Pain is recorded on an ordinal 0-10 scale and accounts for 50% of the overall ASES score. Function accounts for the other 50% of the overall score and is divided into ten questions regarding difficulty with putting on a coat, sleeping on the affected side, washing the back or putting on a bra, managing toileting, combing hair, reaching a high shelf, lifting ten pounds above the shoulder, throwing a ball overhead, participating in work, and participating in sports.

Part 2:

Outcome measures include generic measures, condition-specific measures, and measures of patient satisfaction. The legitimacy and importance of patient satisfaction as an outcome measure has burgeoned over the past decade because of its use for a variety of purposes such as indicating quality of care, assessing health care delivery, developing patient care models, impacting general health status, and allowing for continuous quality improvement. Furthermore, as the paradigm for health care has shifted toward a market model, patient satisfaction has become an outcome measure with great clinical and economic implications.

Part 3:

Expected-value decision analysis is a methodological tool, based in gaming theory, which allows for the quantitative analysis of decision-making under conditions of uncertainty. "Expected-value" refers to the predicted consequences of a decision outcome, which is determined from that outcomes' probability and its utility. Utility is a subjective worth that a patient places on the specific outcome. The process of expected-value decision analysis involves the creation of a decision tree to structure the decision problem, the determination of outcome probabilities and utilities, fold-back analysis to determine the optimal decision-making strategy, and sensitivity analysis to determine the effect of varying outcome probabilities and utilities on decision-making. In this way, decision analysis may allow the clinician and the patient to optimize decision-making based on best-available evidence and patient preferences.

METHODS

Part 1:

Validation of the Lysholm Knee Scale was performed within subsets of a study population of 1,657 patients with chondral disorders of the knee. Lysholm scale, demographic data, subjective assessment, and objective assessment were measured pre-operatively and at periodic post-operative intervals (3 months, 6 months, 12 months, and then yearly), and were maintained prospectively in a computerized database. The mean age of the patients was 44 years old (range: 14-88). There were 1011 (61%) males and 646 (39%) females. Chondral lesions included traumatic chondral injuries involving only one compartment in 679 patients (41%), traumatic chondral injuries involving two or more compartments in 249 patients (15%), and degenerative chondral lesions in 729 patients (44%). Of the 679 patients with traumatic unicompartmental chondral lesions, associated lesions included ligamentous injuries in 230 patients and meniscal injuries in 285 patients. Of the 249 patients with traumatic multicompartmental chondral lesions, associated lesions included ligamentous injuries in 65 patients and meniscal injuries in 107 patients. Of the 729 patients with degenerative chondral lesions, associated lesions included ligamentous injuries in 80 patients and meniscal injuries in 277 patients.

Validation of the Lysholm Knee Scale for meniscal injuries of the knee was performed in subsets of a study population of 191 patients with isolated meniscus pathology. No knee had any ligament pathology or chondral pathology of the knee. Lysholm score, Tegner activity level, demographic data, subjective assessment were measured pre-operatively and at least one year post-operatively and were maintained prospectively in a computerized database. The mean age of 40 years old (range: 13 to 81 years). There were 129 males (67%) males and 62 (33%) females. Medial meniscus pathology was seen in 117 (61%) knees, lateral meniscus pathology in 60 (31%) knees, and medial and lateral meniscus pathology in 14 (7%) knees.

Validation of the ASES Shoulder Score was performed within subsets of a study population. 1,066 patients with instability (n=455), rotator cuff disease (n=474), and arthritis (n=137). Instability patients (n=455) included patients with anterior instability (n=293), multidirectional instability (n=101), and posterior instability (n=61). Mean age for instability patients was 30.3 years old (range: 13.2-74.5 years old) and 62% were male (280/455). Rotator cuff disease patients (n=474) included patients with partial thickness cuff tears (n=223) and full-thickness cuff tears (n=251). Mean age for rotator cuff patients was 56.1 years old (range: 17.8-95.1 years old) and 75% were male (355/474). Arthritis patients (n=137) included those with partial thickness chondral loss (n=8) and full thickness chondral loss (n=129) due to either osteoarthritis (n=108), avascular necrosis (n=7), or rheumatoid arthritis (n=22). Mean age for arthritis patients was 62.3 years old (range: 32.9-82.5 years old) and 85% were male (116/137). ASES shoulder scale, demographic data, subjective assessment, and objective assessment were measured pre-operatively and at periodic post-operative intervals (3 months, 6 months, 12 months, and then yearly), and were maintained prospectively in a computerized database.

Test-retest reliability was determined by having patients complete an original preoperative questionnaire and a second preoperative questionnaire within four weeks of the original questionnaire. There was no interval change in health status as ascertained by health history forms with a complete review of systems. The intraclass correlation coefficient was determined for the overall scale and for subscales. An intraclass correlation coefficient >0.70 was considered acceptable. Internal consistency was determined using preoperative scores. Overall internal consistency was determined. Cronbach's alpha >0.60 was considered acceptable. Content validity was determined using preoperative scores. Preoperative Lysholm knee scales were used to establish content validity. Floor effects (scale= lowest possible) and ceiling effects (scale=highest possible) were determined for the overall score and for subscales. Floor and ceiling effects <30% were considered acceptable. Criterion validity was determined by correlations to accepted outcome instruments, the physical functioning, role physical, and bodily pain domains of the SF12 or the SF36 and the pain, stiffness, and function domains of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). The Pearson correlation coefficient was used for the continuous outcome measures (SF12 and Tegner), whereas Spearman's rho was used for categorical outcome measure (WOMAC). Construct validity was determined by constructing multiple noncontroversial hypotheses. These hypotheses were then tested using appropriate statistical testing of the outcome score. Responsiveness to change was assessed by comparing preoperative scores compared to postoperative scores after effective treatment. Effect size was calculated: (mean postoperative scale – mean preoperative scale) / standard deviation of preoperative scale. Standardized response mean was calculated: (mean postoperative scale – mean preoperative scale) / standard deviation of the change in scale. Small effects were considered >0.20, moderate effects were considered >0.50, and large effects were considered >0.80.

Part 2:

Patient satisfaction with outcome after ACL reconstruction was performed in a cohort of patients (n=201) undergoing primary anterior cruciate ligament reconstruction. All patients included in this cohort had two-year minimum follow-up (mean=35.9 months; range: 24-87 months) with complete preoperative, surgical, and post-operative subjective and objective data. Mean age of the cohort was 28.6 years (range: 14.4-60.0 years old) and 56.9% of the cohort was male. Exclusion criteria included contralateral anterior cruciate ligament injury and ipsilateral revision reconstruction.

Patient satisfaction with outcome after rotator cuff surgery was performed in a cohort of patients (n=311) undergoing rotator cuff surgery from 1991-1998. Two hundred fifty-four patients had minimum two-year subjective follow-up (mean=33.7 months; range, 24-89 months) and 57 patients had minimum one-year subjective and objective follow-up (mean=30.7 months; range, 24-71 months). Mean age of the cohort was 51.4 years old (Standard Deviation: 14.2: range, 16.5-80.3) and 69.5% (216/311) of the cohort was male. Exclusion criteria included prior rotator cuff surgery, contralateral surgery, lack of minimum one-year subjective or objective follow-up, and patients undergoing tendon transfer or tendon grafting. Diagnoses included tendinitis/impingement (n=80), partial thickness tear (n=55), and full thickness tear (n=176). Patients with impingement (n=80) underwent arthroscopic subacromial decompression. Patients with partial thickness tears (n=55) underwent either arthroscopic subacromial decompression (n=45) or arthroscopic subacromial decompression with open repair through a deltoid-splitting mini-open approach for high-grade partial tears (n=10). Patients with full thickness tears (n=176) underwent arthroscopic subacromial decompression with either open repair through a deltoid-splitting mini-open approach (n=159) or debridement for a massive, irreparable tear (n=17). Distal clavicle excision was performed in 29 patients (20 arthroscopic, 9 open).

The dependent variable was patient satisfaction with outcome at the time of follow-up. This was assessed on a scannable paper questionnaire administered in the office waiting room, asking, "How satisfied are you with your outcome?"

This question was scored on a discrete ordinal scale from 1-10, framed by "very satisfied" at 10, "neutral" between 5 and 6, and "very unsatisfied" at 1. Independent variables included various demographic, surgical, objective at follow-up, and subjective at follow-up parameters.

Nonparametric univariate analysis was performed with the Mann-Whitney U test for two group comparisons and the Kruskal-Wallis ANOVA for multiple group comparisons. Spearman's rho (r) was used for assessing associations between continuous variables. Multivariate analysis was performed to identify independent determinants of patient satisfaction with outcome using a stepwise, multiple linear regression model with backward selection. Collinearity and residual diagnostics were performed. The coefficient of multiple determination (adjusted R2) was used to indicate how much of the variability in patient satisfaction was accounted for by the determinants in the final multiple linear regression model.

Part 3:

The process of expected-value decision analysis involves the creation of a decision tree to structure the decision problem, the determination of outcome probabilities and utilities, fold-back analysis to determine the optimal decision-making strategy, and sensitivity analysis to determine the effect of varying outcome probabilities and utilities on decision-making. Outcome probabilities were determined from quantitative literature syntheses of Achilles tendon ruptures and SCFE. Outcome utilities were obtained for using a visual-analog scale questionnaire given to patients with Achilles tendon rupture. Questionnaires posed scenarios for the different outcomes and asked subjects to rate these outcomes on a zero to ten scale, where zero represented the worst possible medical outcome and ten represented the best possible medical outcome. Further questions or clarifications about the disease state were answered by the investigators. Decision trees were constructed with decision branches (decision node), chance branches (chance nodes), and outcomes (terminal nodes). Per convention, outcome utilities were placed to the right of the terminal nodes and outcome probabilities were placed under the terminal nodes. Fold-back analysis was performed by determining the weighted-average of outcome probabilities multiplied by outcome utilities to determine the overall expected value. The expected values were then compared to determine the optimal decision. One-way sensitivity analysis was performed to model the effect of varying a probability or a utility. Two-way sensitivity analysis was performed to model the effect of varying two variables simultaneously.

Table 2: Floor-Ceiling Effects

Domain	Floor Effects	Ceiling Effects
Overall Lysholm scale (0-100)	0 (0%)	12 (0.7%)
Pain (0-25)	430 (26%)	74 (4.5%)
Swelling (0-10)	201 (12.1%)	318 (19.2%)
Limp (0-5)	145 (8.8%)	537 (32.4%)*
Squatting (0-5)	859 (51.8%)*	99 (6.0%)
Instability (0-25)	83 (5.0%)	1104 (66.6%)*
Support (0-5)	226 (13.6%)	914 (55.2%)*
Stair Climbing (0-10)	323 (19.5%)	211 (12.7%)
Locking (0-15)	16 (1.0%)	1023 (61.7%)*

* >30% floor or ceiling effect

Table 3: Criterion Validity

Scale	Domain	Correlation Coefficient	P Value
SF12	Physical Functioning	0.537*	<0.001
	Role Physical	0.476*	<0.001
	Bodily Pain	0.559*	<0.001
WOMAC	Pain	-0.802#	<0.001
	Stiffness	-0.657#	<0.001
	Function	-0.814#	<0.001
Tegner		0.346*	<0.001

*Pearson correlation coefficient, #Spearman's rho

Table 1: Test-Retest Reliability

Component	Intraclass Correlation Coefficient
Overall Lysholm scale	0.91 (0.82-0.98)
Pain	0.61* (0.50-0.77)
Instability	0.82 (0.73-0.91)
Locking	0.97 (0.90-0.99)
Stair-Climbing	0.67* (0.53-0.83)
Limping	0.82 (0.71-0.92)
Support	0.98 (0.91-1.00)
Swelling	0.94 (0.85-0.99)
Squatting	0.91 (0.82-0.98)

Values represent coefficient and 95% confidence intervals

*intraclass correlation coefficient <0.70.

RESULTS

Part 1:

For the validation of the Lysholm Knee Score for chondral disorders, test-retest reliability for is shown in Table 1. There was acceptable (Cronbach's alpha >0.60) internal consistency for the Lysholm scale (Cronbach's alpha= 0.65). Floor and ceiling effects are shown in Table 2. Criterion validity results are shown in Table 3. For construct validity, all nine hypotheses (constructs) were significant (P<0.05).

Table 4: Responsiveness

Domain	Effect Size	Standardized Response Mean
Overall Lysholm scale*	1.16	1.10
Pain*	1.31	1.28
Instability+	0.21	0.20
Locking#	0.55	0.50
Stair-Climbing#	0.75	0.70
Limping*	1.29	1.27
Support#	0.59	0.54
Swelling*	1.17	1.08
Squatting*	1.25	1.24

*large effect (>0.80), #moderate effect (>0.50), +small effect (>0.20)

Table 5: Test-Retest Reliability

Component	Intraclass Correlation Coefficient
Overall Lysholm scale	0.927 (0.90-0.95)
Pain	0.73 (0.64-0.81)
Instability	0.87 (0.82-0.91)
Locking	0.67* (0.56-0.76)
Stair-Climbing	0.81 (0.74-0.87)
Limping	0.76 (0.67-0.83)
Support	**
Swelling	0.75 (0.66-0.82)
Squatting	0.86 (0.81-0.90)

Values represent coefficient and 95% confidence intervals, *Intraclass correlation coefficient <0.70., **No variance in item.

Table 6: Floor-Ceiling Effects

Domain	Floor Effects		Ceiling Effects	
	Group A	Group C	Group A	Group C
Overall Lysholm scale (0-100)	0(0%)	0(0%)	1(0.5%)	2(0.4%)
Pain (0-25)	55(28%)	86(17%)	10(5.2%)	24(5%)
Swelling (0-10)	14(7.4%)	51(10%)	43(22.5%)	161(33%)*
Limp (0-5)	23(12%)	57(12%)	70(37%)*	148(30%)
Squatting (0-5)	64(34%)*	71(14%)	17(9%)	72(14%)
Instability (0-25)	8(4%)	12(2%)	106(55%)*	129(26%)
Support (0-5)	10(5%)	9(2%)	167(87%)*	434(88%)*
Stair Climbing (0-10)	19(10%)	6(1.2%)	43(22%)	138(28%)
Locking (0-15)	7(4%)	7(1.4%)	98(51%)*	204(41%)*
Tegner (Group A)	2(2.5%)		2(2.5%)	
Tegner (Group C)	26(5.2%)		0(0%)	

* >30% floor or ceiling effect

Table 7: Criterion Validity of the Lysholm Score

Scale	Domain	Correlation Coefficient	P Value
SF12	Physical Functioning	0.541*	<0.001
	Role Physical	0.420*	<0.001
	Bodily Pain	0.551*	<0.001

*Pearson correlation coefficient

1. Patients with lower activity levels had significantly lower Lysholm scales (n=1657, r=0.410; P<0.001).
2. Patients with a greater number of chondral surfaces with Outerbridge grade 4 changes had significantly lower Lysholm scales (F=4.73, P<0.001) (0 surfaces: n=776, mean=59.9, standard deviation=19.0; 1 surface: n=396, mean=59.1, standard deviation=19.3; 2 surfaces: n=373, mean=55.3, standard deviation =18.5; 3 surfaces: n=63, mean=55.4, standard deviation=16.5; 4 surfaces: n=34, mean=54.9, standard deviation=19.6; 5 surfaces: n=4, mean =49.3, standard deviation=10.6; 6 surfaces: n=11, mean=42.6, standard deviation=26.2).
3. Patients with full thickness chondral defects (Outerbridge grade 4) had significantly lower Lysholm scales than patients with partial thickness chondral defects (Outerbridge grade 3) (full-thickness defects: n=881, mean=56.7, standard deviation=19.0; partial-thickness defects: n=776, mean=59.9, standard deviation=19.0; P=0.001).
4. Patients with chondral defects and associated meniscal tears had significantly lower Lysholm scales that patients with isolated chondral defects (chondral + meniscus: n=795, mean=56.4, standard deviation=19.2; chondral alone: n=862, mean=59.1, standard deviation=18.8; P=0.01).
5. Patients with more difficulty with activities of daily living had significantly lower Lysholm scales that patients with less difficulty with activities of daily living (n=1657, r=0.421, P<0.001).
6. Patients with more difficulty working because of their knee had significantly lower Lysholm scales that patients with less difficulty working because of their knee (n=1657, r=0.407, P<0.001).
7. Patients with more difficulty with sports because of their knee had significantly lower Lysholm scales that patients with less difficulty with sports because of their knee (n=1657, r=0.330, P<0.001).
8. Patients with previous knee surgery had significantly lower Lysholm scales than patients without previous knee surgery (previous surgery: n=848, mean=56.7, standard deviation=18.9; no previous surgery: n=774, mean=59.9, standard deviation=19.0; P=0.001).

9. Patients with a poorer assessment of overall knee function had significantly lower Lysholm scales that patients with a better assessment of overall knee function (n=1657, r=0.475, P<0.001).

Responsiveness to change results are shown in Table 4.

For the validation of the Lysholm Knee Score for meniscal injuries, test-retest reliability is shown in Table 5. There was acceptable (Cronbach's alpha >0.60) internal consistency for the Lysholm score (Cronbach's alpha= 0.729). Floor and ceiling effects are shown in Table 6. Criterion validity results are shown in Table 7. For construct validity, all eight hypotheses (constructs) were significant (P<0.05).

1. Patients with lower activity levels had lower Lysholm scores. (r=0.550, p<0.001)
2. Patients with acute injury had lower Lysholm scores than patients with a chronic knee injury. (Acute patients mean Lysholm = 51.6, standard deviation =20.1; chronic patients mean Lysholm = 60.3, standard deviation=19.2, p= 0.012)
3. Patients with a worker's compensation claim had lower Lysholm scores than patients without a worker's compensation claim. (Worker's compensation mean Lysholm = 47.1, standard deviation =18.9; non-worker's compensation mean = 59.4, standard deviation=22.5, p=0.015)
4. Patients with more difficulty with activities of daily living had lower Lysholm scores than patients with less difficulty with activities of daily living. (r=0.508, p<0.001)
5. Patients with more difficulty working because of their knee had lower Lysholm scores than patients with less difficulty working because of their knee. (r=0.528, p<0.001)
6. Patients with more difficulty with sports because of their knee had lower Lysholm scores than patients with less difficulty with sports because of their knee. (r=0.375, p<0.001)
7. Patients with abnormal or severely abnormal assessment of overall knee function had lower Lysholm scores than patients with a normal or nearly normal assessment of overall knee function. (Abnormal knee function mean Lysholm =49.9, standard deviation=17.7; normal knee function mean Lysholm = 72.0, standard deviation=14.5 , p<0.001)
8. Patients with degenerative/complex meniscus tears had lower Lysholm scores than patients with simple tears of the meniscus. (Degenerative/complex tears mean Lysholm score =55.4, standard deviation=18.8; simple tears mean Lysholm score=63.9, standard deviation=19.4, p= 0.004)

Table 8: Responsiveness

Domain	Effect Size		Standardized Response Mean	
	Group A	Group C	Group A	Group C
Overall Lysholm scale	1.2*	1.2*	0.97*	1.13*
Pain	1.2*	0.97*	1.05*	0.88*
Instability	0.15	0.89*	0.10	0.84*
Locking	0.50#	0.69#	0.50#	0.61#
Stair-Climbing	0.89*	0.87*	0.77	0.78
Limping	0.69#	0.74#	0.52#	0.72#
Support	0.21+	0.21+	0.18	0.18
Swelling	0.70#	0.63#	0.70#	0.57#
Squatting	1.04*	0.66	0.86*	0.60

*large effect (>0.80), #moderate effect (>0.50), +small effect (>0.20)

Table 9: Test-Retest Reliability

Component	Intraclass Correlation Coefficient
Overall ASES shoulder scale	0.94 (0.88-0.97)
Pain	0.83 (0.69-0.91)
Do usual work	0.90 (0.80-0.95)
Do usual sport	0.88 (0.77-0.94)
Put on a coat	0.85 (0.73-0.93)
Sleep on affected side	0.71 (0.49-0.83)
Wash back	0.92 (0.85-0.96)
Manage toileting	1.00 (1.00-1.00)
Comb hair	0.75 (0.55-0.86)
Reach a high shelf	0.93 (0.87-0.96)
Lift 10 pounds above shoulder	0.82 (0.68-0.90)
Throw a ball overhead	0.76 (0.58-0.88)

Values represent coefficient and 95% confidence intervals

Table 10: Floor-Ceiling Effects

Population	Floor Effects	Ceiling Effects
Shoulder Instability	0% (0/455)	1.3% (6/455)
Rotator Cuff Disease	0% (0/474)	0% (0/474)
Glenohumeral Arthritis	0% (0/137)	0% (0/137)

For the validation of the ASES Shoulder Score, test-retest reliability for is shown in Table 9. There was acceptable (Cronbach's alpha >0.60) internal consistency for the ASES shoulder scale in patients with shoulder instability (Cronbach's alpha= 0.61), rotator cuff disease (Cronbach's alpha= 0.64), and glenohumeral arthritis (Cronbach's alpha= 0.62). Floor and ceiling effects are shown in Table 10. Criterion validity results are shown in Table 11. Responsiveness is shown in Table 12. Construct validity is shown in Table 13 ($P<0.05$).

Part 2:

Multivariate predictors of patient satisfaction with outcome after ACL reconstruction are shown in Table 14.

Multivariate predictors of patient satisfaction with outcome after rotator cuff surgery are shown in Table 15.

Part 3:

For expected-value decision analysis for Achilles Tendon rupture, the decision tree is shown in Figure 16. Operative

Table 11: Criterion Validity

Scale	Domain	Correlation Coefficient	P Value
SF12	Physical Functioning	0.57	<0.001*
	Role Physical	0.32	0.002*
	Bodily Pain	0.58	<0.001*
	Role Emotional	-0.09	0.49
	Mental Health	-0.08	0.67
	Vitality	0.11	0.27
	Social Function	0.10	0.34

*P<0.05

Table 12: Responsiveness

Population	Pre-Operative Score	Post-Operative Score	Change Score	Effect Size	Standardized Response Mean
Shoulder Instability	64.0 + 22.7	83.6 + 21.4	17.9 + 21.1	0.86	0.93
Rotator Cuff Disease	51.3 + 18.6	83.7 + 22.0	29.2 + 20.1	1.33	1.16
Glenohumeral Arthritis	56.5 + 19.3	82.1 + 20.9	24.7 + 22.2	1.74	1.11

Table 13: Construct Validity

#	Hypothesis	Scale of Independent Variable	Statistical Test	P value
Shoulder Instability				
1	Patients with less satisfaction with their shoulder function would have lower ASES shoulder scale scores.	Ordinal 0-10 scale, anchored at completely dissatisfied (0) and completely satisfied (10).	Pearson correlation coefficient	<0.001*
2	Worker's compensation patients would have lower ASES shoulder scale scores than non-worker's compensation patients.	Dichotomous (worker's compensation, non-worker's compensation)	Student's t-test	<0.001*
3	Patients with greater shoulder-related difficulty with activities of daily living would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.008*
4	Patients with greater shoulder-related difficulty with work would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.003*
5	Patients with greater shoulder-related difficulty with sports would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.018*
6	Patients with greater shoulder-related difficulty with sleep would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.001*
7	Patients with greater frequency of shoulder instability would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, rare, occasional, and frequent.	Spearman's rho	0.016*
8	Patients with less forward elevation would have lower ASES shoulder scale scores.	Ordinal Likert scale of waist, nipple, neck, head, and overhead.	Spearman's rho	<0.001*
Rotator Cuff Disease				
9	Patients with less satisfaction with their shoulder function would have lower ASES shoulder scale scores.	Ordinal 0-10 scale, anchored at completely dissatisfied (0) and completely satisfied (10).	Pearson correlation coefficient	<0.001*
10	Worker's compensation patients would have lower ASES shoulder scale scores than non-worker's compensation patients.	Dichotomous (worker's compensation, non-worker's compensation)	Student's t-test	<0.001*
11	Patients with greater shoulder-related difficulty with activities of daily living would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	<0.001*
12	Patients with greater shoulder-related difficulty with work would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.001*

13	Patients with greater shoulder-related difficulty with sports would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.001*
14	Patients with greater shoulder-related difficulty with sleep would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	<0.001*
15	Patients with greater difficulty carrying 20 pounds at their side would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.048*
Glenohumeral Arthritis				
16	Patients with less satisfaction with their shoulder function would have lower ASES shoulder scale scores.	Ordinal 0-10 scale, anchored at completely dissatisfied (0) and completely satisfied (10).	Pearson correlation coefficient	<0.001*
17	Patients with greater shoulder-related difficulty with activities of daily living would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.041*
18	Patients with greater shoulder-related difficulty with sports would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.027*
19	Patients with greater shoulder-related difficulty with sleep would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.038*
20	Patients with greater difficulty tucking in their shirt would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.034*
21	Patients with greater difficulty lifting one pound to shoulder height would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	<0.001*
22	Patients with greater difficulty carrying 20 pounds at their side would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.001*
23	Patients with greater difficulty washing their back or opposite shoulder would have lower ASES shoulder scale scores.	Ordinal Likert scale of none, mild, moderate, and severe.	Spearman's rho	0.040*

*P<0.05

Table 14

Variable	Scale	Regression Coefficient (β)	Standard Error	P Value
Lysholm Score	0-100	0.182	0.012	0.049
Overall Knee Function	Normal, Nearly Normal, Abnormal, Severely Abnormal	-0.327	0.232	<0.001
IKDC Range of Motion Subscore	A=Normal, B=Nearly Normal, C=Abnormal, D=Severely Abnormal	-0.247	0.367	<0.001
Patella Tenderness	None, Mild, Moderate, Severe	-0.277	0.452	<0.001
Full Giving-Way	None, Mild, Moderate, Severe	-0.315	0.283	<0.001
Flexion Contracture	No, Yes (>5° contracture)	-0.157	0.426	0.010
Swelling	None, Mild, Moderate, Severe	-0.129	1.096	0.026

Table 15

Variable	Scale	Regression Coefficient (B)	Standard Error	p Value
Pain Today	0-10	-0.44	0.17	<0.001
Return to Work	No, Yes	0.31	0.49	0.022
Dysfunction with Toileting	1-4	-0.29	0.63	0.046
Dysfunction with 20 lbs	1-4	-0.19	0.40	0.047
Recommend Surgery?	No, Yes	0.18	0.16	0.034
Active Forward Elevation	Degrees	0.82	0.57	0.048
ASES Score	0-100	1.20	0.09	0.046

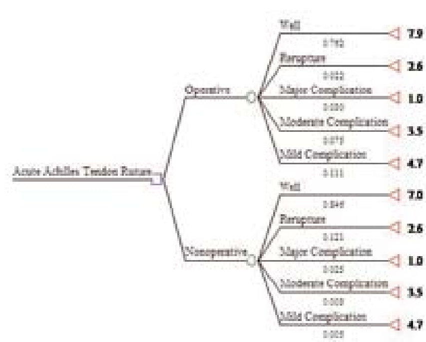


Figure 16

management was favored. Sensitivity analysis for the probability of a moderate complication (mainly wound complication) is shown in Figure 17.

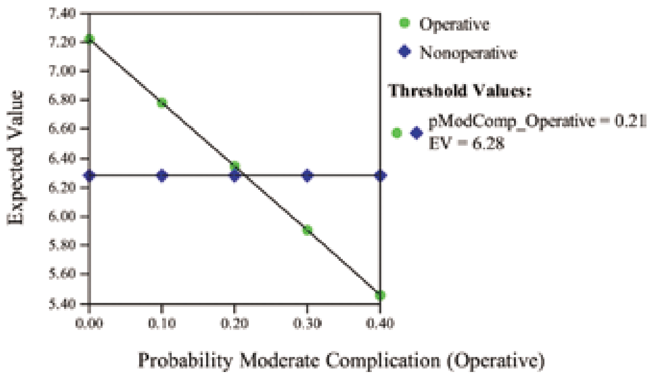


Figure 17

For expected-value decision analysis for the contralateral hip in SCFE, the decision tree is shown in Figure 18. Sensitivity analysis for the probabililty of a contralateral slip and for the utility of pinning is shown in Figure 19.

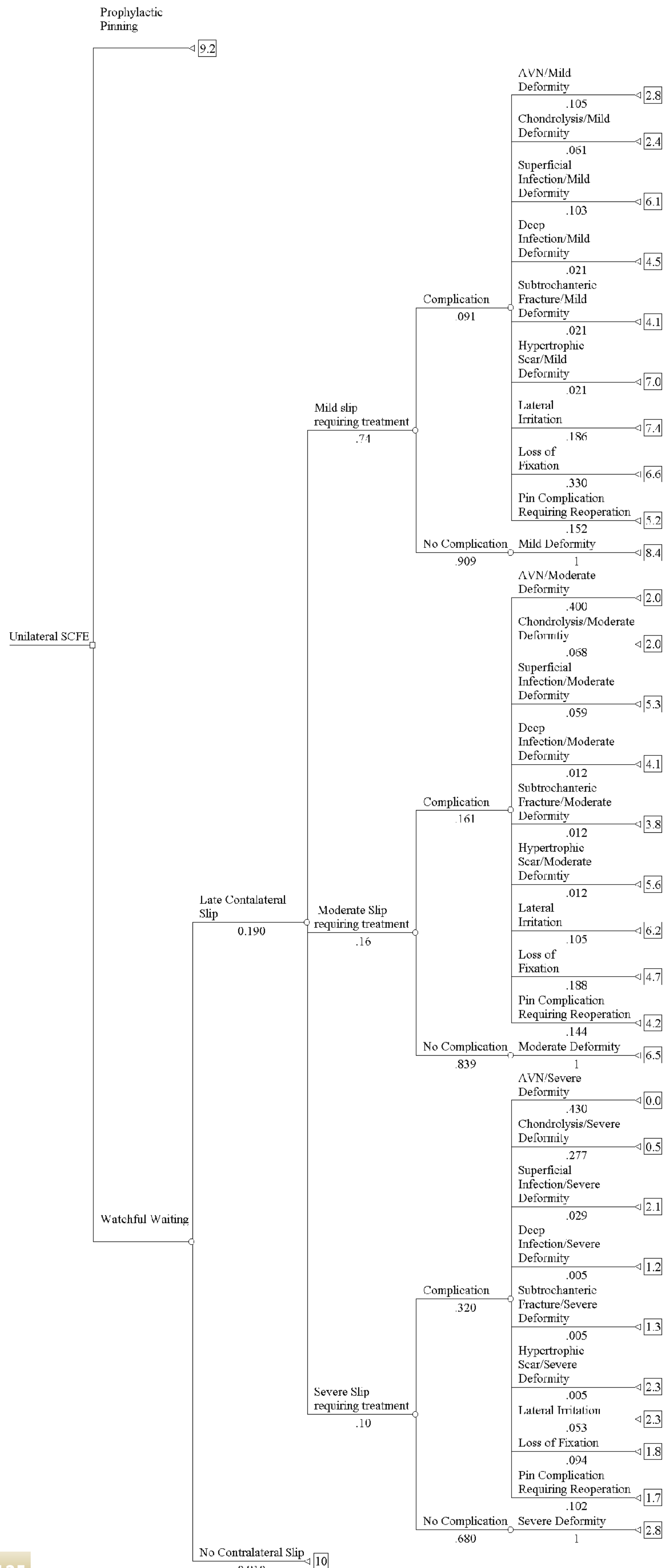
DISCUSSION

Outcomes research has become the pivotal means to assessment and accountability in medicine. The emphasis of the outcomes assessment movement has been patient-derived outcomes assessment. The goal of these studies was to emphasize patient-derived outcomes assessment through the validation of outcomes instruments, the study of patient satisfaction, and the utilization of patient preferences in expected-value decision analysis.

In Part 1, we validated the Lysholm Knee Score for chondral and meniscal disorders of the knee and the ASES Shoulder Score for various disorders of the shoulder. The Lysholm knee scale and the ASES Shoulder score demonstrated, in general, acceptable psychometric parameters (test-retest reliability, internal consistency, floor-ceiling effects, criterion validity, construct validity, and responsiveness) to justify their use in outcomes assessment for various chondral and meniscal disorders of the knee and various shoulder disorders.

In Part 2, we determined the predictors of patient satisfaction after ACL reconstruction. We found that some specific surgical and objective variables were associated with patient satisfaction with outcome after anterior cruciate ligament reconstruction, however, the most robust associations were with subjective variables of symptoms and function. Clinically relevant groupings of variables affecting satisfaction included issues of stiffness, giving-way, swelling, and patellofemoral symptoms. Thus, in assessing the outcomes of anterior cruciate ligament reconstruction from the perspective of patient satisfaction with outcome, we would emphasize the importance of patient-derived subjective assessment of symptoms and function, consistent with the general philosophy of the health care outcomes assessment movement. In particular, we would focus on the parameters involving issues of stiffness, giving-way, swelling, and patellofemoral symptoms. For rotator cuff surgery, we found as with ACL reconstruction, that while some

Figure 18



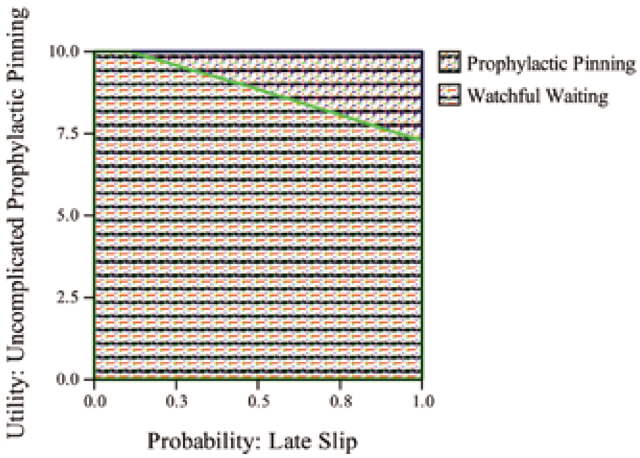


Figure 19

specific surgical and objective variables were associated with patient satisfaction after rotator cuff surgery, the most robust associations were with subjective variables of symptoms and function.

In part 3, we utilized expected-value decision analysis to explore decision making from the patient's perspective for Achilles tendon rupture and management of the contralateral hip in SCFE. Using outcome probablilities from a quantitative literature synthesis and utilities from patient preferences, we determined that the optimal management for Achilles tendon rupture was operative and the optimal management of the contralateral hip in SCFE was observation.

References

1. **Relman AS:** Assessment and accountability: The third revolution in medical care. *N Engl J Med.* 1988;319:1220-2.
2. **Kane RL:** Outcome measures. In Understanding Health Care Outcomes Research, pp. 17-18. Edited by Kane R, 17-18, Gaithersberg, Aspen Publishers, 1997.
3. **Kocher MS, Steadman JR, Briggs KK, Sterett WI, Hawkins RJ.** Reliability, validity, and responsiveness of the Lysholm knee scale for chondral disorders of the knee. J Bone Joint Surg, 2004, 86-A(6): 1139-1145.
4. **Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ.** Reliability, validity, and responsiveness of the American shoulder and elbow surgeons shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg, 2005..
5. **Briggs KK, Kocher MS, Rodkey W, Steadman JR.** Reliability, validity, and responsiveness of the Lysholm knee score and Tegner activity scale for meniscus injuries of the knee. J Bone Joint Surg, 2005.
6. **Kocher MS, Steadman JR, Zurakowski D, Briggs K, Sterett WI, Hawkins RJ.** Determinants of patient satisfaction after anterior cruciate ligament reconstruction. J Bone Joint Surg, 2002, 84-A(9): 1560-1572.
7. **O'Holleran JD, Kocher MS, Horan MP, Briggs KK, Hawkins RJ.** Determinants of patient satisfaction with outcome after rotator cuff surgery. J Bone Joint Surg, 2005, 87-A(1):121-6.
8. **Kocher MS, Bishop J, Luke A, Marshall R, Hawkins RJ.** Operative vs Nonoperative Management of Acute Achilles Tendon Ruptures: Expected-Value Decision Analysis. Am J Sports Med, 2002. 30(6): 783-790.
9. **Kocher MS, Bishop JB, Hresko MT, Kim YJ, Millis MB, Kasser JR.** Prophylactic pinning of the contralateral hip after unilateral slipped capital femoral epiphysis: expected-value decision analysis. J Bone Joint Surg, 2004, 86-A(12):2658-65.