Validity, Reliability, and Responsiveness of the Anterior Cruciate Ligament Quality of Life Measure: A Continuation of Its Overall Validation

Mark R. Lafave, PhD,† Laurie Hiemstra, MD, PhD,† Sarah Kerslake, MSc, BPhty,† Mark Heard, MD,† and Greg Buchko, MD†*

Objective: The purpose is to provide more validity, reliability, and responsiveness testing of the anterior cruciate ligament–quality of life instrument (ACL-QOL), particularly in light of consensus-based standards for the selection of health status measurement instruments (COSMIN) guidelines.

Design: Prospective case series.

Setting: An orthopedic surgical practice for consultation.

Patients: A convenience sample of 579 ACL-deficient patients.

Intervention: Anterior cruciate ligament reconstructive surgery.

Main Outcome Measures: Patients completed the ACL-QOL at initial visit and underwent reconstructive surgery. Patients were followed at 6, 12, and 24 months using the ACL-QOL to determine its validity and responsiveness. Cronbach's alpha was used to determine the unidimensionality of the ACL-QOL. A subset of patients took the ACL-QOL twice in a test–retest reliability analysis (intraclass correlation coefficient or ICC 2,k). Another subset of 24-month postsurgical patients measured the success of their surgery using a 7-point global rating scale of improvement as an anchor-based method of responsiveness.

Results: Cronbach's alpha coefficients = 0.93, 0.95, 0.96, and 0.98 at 6, 12, and 24 months, respectively. Intraclass correlation coefficient = 0.60, SEM = 6.16, and confidence interval of 12.1 (CI 95%). Responsiveness was measured by comparing the 4 serial time periods. Patients improved significantly at each time period ($P < 0.05$, ETA squared 0.61). A 24-month ACL-QOL was significantly correlated ($P > 0.05$) to being "significantly better" or "somewhat better."

Conclusions: The results of this study added more validity, reliability, and responsiveness for the ACL-QOL. The ACL-QOL has completed 8 of 9 COSMIN criteria.

Key Words: psychometric, clinimetric, knee, quality of life

(*Clin J Sport Med* 2017;27:57–63)

Submitted for publication May 24, 2015; accepted September 23, 2015.
From the *Mount Royal University, Calgary, Alberta, Canada; and †Banff Sport Medicine, Banff, Alberta, Canada.
The authors report no conflicts of interest.
Corresponding Author: Mark R. Lafave, PhD, Mount Royal University, 4825 Mount Royal Gate SW, Calgary, AB, T3E 6K6 (mlafave@mtroyal.ca).

INTRODUCTION

There are a number of patient-reported and objective measurements that can be used when making decisions about the success of treatment of anterior cruciate ligament deficient knees. When the anterior cruciate ligament–quality of life (ACL-QOL) instrument was first introduced in 1995, disease-specific, QOL, patient-reported outcome measures (PROMs) were not as prevalent in the literature as they are today.[1,2] The initial validity and reliability of the ACL-QOL for assessing ACL-deficient patients were published in 1998.[1] The ACL-QOL is a 32-item patient-reported questionnaire scored using a 100-mm visual analog scale (VAS) and converted to an overall score out of 100, where each question receives equal weight.[1] The domains that comprise the ACL-QOL are (1) symptoms and physical complaints (5 items), (2) work-related concerns (4 items), (3) recreational activity and sport participation or competition (12 items), (4) lifestyle (6 items), and (5) social and emotional (5 items). It is important to note that there is a combination of 3 features of the ACL-QOL that makes it unique: (1) it is disease-specific, (2) it is a PROM, and (3) it is a QOL measure.

There have been a number of studies that have attempted to address the issue of the "best" outcome measure, particularly as it relates to the knee.[3–6] In light of the 3 variables mentioned previously [ie, (1) disease-specificity, (2) PROM, and (3) QOL], passing judgment about the best outcome measure is more complicated than the binary question posed. Some outcome measure instruments gather patient data generically for knee conditions, but not specific to the ACL.[7–9] Other instruments measure QOL for diseases such as osteoarthritis,[10,11] and some instruments measure QOL generically, not specific to any disease or condition.[12] Several instruments measure only 1 or 2 dimensions of QOL, such as activity level or symptoms for patients with knee injuries,[13,14] whereas other instruments measure deficits in ACL-injured patients through more "objective" means such as functional testing.[5,15] Therefore, it is critical that the question of the best outcome measure be qualified with specific criteria.

The ACL-QOL is very specific for an ACL-injured population and the questions are highly valued by the majority (51%) of patients when compared with 11 knee-evaluation instruments.[4] When judging the best instrument to evaluate knee or ACL-injuries, the ACL-QOL has consistently demonstrated good content validity.[1,4] Garratt et al[3] compared 16 PROM for knee injuries on 3 measures: validity, reliability, and responsiveness. Specific to the ACL-QOL, they identified 2 factors missing from the initial ACL-QOL publication: internal consistency and responsiveness. Absence of these variables does not inherently render the instrument invalid but rather demonstrates a need for further assessment of these factors. Barber-Westin et al[16] were critical of the ACL-QOL because it solely relied on patient-perceived variables and not more objective measures such as anterior-posterior displacement, radiographic analysis, range of motion testing or functional testing. These assertions are accurate, particularly as the primary objective of the ACL-QOL is to be a patient-reported, health-related QOL assessment that is specific to 1 condition: ACL injury. Patient-reported disease-specific (ACL injury) evaluation instruments, as opposed to generic QOL measures, have been proven to provide more valid, reliable, and responsive measurements for patients across a number of diseases or conditions.[17–19] Although the high number of assessment instruments for the knee makes it confusing for clinicians to pick the best instrument, it is important to use an instrument that measures the outcomes that are most important to patients, while also supplementing outcome measures that are important to clinicians. Therefore, if there are certain psychometric or clinimetric measures missing from the ACL-QOL, further development and evaluation are warranted.

A QOL measure assesses a person as a whole and as such it is intended to measure a person's "health state."[19] Quality of life has been defined as a concept encompassing a broad range of physical and psychological characteristics and limitations, which describe an individual's ability to function and derive satisfaction from doing so.[20] Quality of life measures can assist health care practitioners in making sound evidence-based clinical decisions and have become the primary outcome measure for clinical decision-making.[5] For example, Naylor et al[21] used the ACL-QOL to evaluate the efficacy of preserving the ACL in reconstructive surgery. These authors concluded, based on patient responses at the 1-year period, that preservation of the ACL resulted in improved QOL outcomes compared with those patients who did not receive ACL preservation. Such hypothesis testing using a QOL measure that is also disease-specific is critical for clinician decision making. Quality of life measures can also facilitate economic decisions regarding the best type of care. Quality of life measures are typically used to calculate the quality-adjusted life year, which can be used to evaluate the efficacy of various treatments.[19] Therefore, QOL measures are quite versatile in their utility, assisting from both a clinical decision-making process assessing the best treatments, while also factoring in economic and policy implications.[19]

The consensus-based standards for the selection of health status measurement instruments (COSMIN) were introduced as taxonomy to assess measurement properties for health-related QOL, PROMs.[22] One goal of developing these standards was to provide guidelines to assist practitioners in choosing generic or disease-specific QOL measures in their clinical practice. The guidelines were also intended to guide the evaluation of studies in systematic reviews. The COSMIN taxonomy has been content validated[22] and demonstrated reliability.[23] The COSMIN guidelines have been criticized as not being designed to assess the quality of an instrument.[24] Despite this criticism, the COSMIN checklist does provide an excellent framework to analyze PROMs such as the ACL-QOL.[1]

PURPOSE

The purpose of this article is to evaluate the ACL-QOL in the context of the COSMIN guidelines to provide additional validity, reliability, and responsiveness testing of the ACL-QOL. In addition, the purpose of this article was to place the results into historical context by inventorying the psychometric and clinimetric features of the ACL-QOL from the literature that has used it.

METHODS

Patient Population

A convenience sample was used in this study for prospective analysis. Five hundred seventy-nine ACL-deficient patients were referred to an orthopedic surgical practice for consultation. All patients completed the ACL-QOL preoperatively as a baseline measurement.[1] All patients underwent single-bundle anatomic ACL reconstruction with a hamstring tendon graft. Surgery was performed by 1 of 3 fellowship trained orthopedic surgeons, each of whom perform more than 200 ACL reconstruction procedures annually (L.H., M.H., G.B.). None of these patients underwent additional major procedures such as osteotomy or multiple ligament reconstruction. This sample did include patients who required a partial meniscectomy or meniscal repair. These patients are from a prospective cohort that is being followed from initial consultation at 6, 12, and 24 months postoperatively.

Ethical Considerations

Patients who were referred to the Banff Sport Medicine clinic all complete the ACL-QOL as part of normal practice. All patients who met the inclusion criteria for the study were asked if they wanted to volunteer to participate in this study. The University of Calgary Conjoint Health Research Ethics Board approved this study.

Content Validity Testing

To test the unidimensionality of the ACL-QOL, Cronbach's alpha was calculated at the initial consultation, and at 6, 12, and 24 months postoperatively.[25] Subjects were stratified based on their ACL-QOL scores and separated to assess for floor or ceiling effects. Floor and ceiling effects were calculated at the 15%, 20%, 25%, and 30% stratifications for each time period.

Reliability Testing

A convenience sample (n = 31) of the larger sample (n = 579) of patients was recruited to measure the reliability of the ACL-QOL. The intraclass correlation coefficient (ICC) is a measure of variance between trials, thus relative reliability. The ACL-QOL was sent to a random sample of

50 patients and asked to respond again 2 weeks later. The questionnaires from those patients who completed the ACL-QOL within the 2-week time frame were used in the statistical analysis. A Cronbach's alpha and an ICC (2,k) was used to measure the reliability coefficient of the ACL-QOL. The SEM was calculated as a measure of stability by multiplying the square root of 1 minus the ICC by the SD of the baseline scores of the ACL-QOL.[26,27] The SEM is a measure of absolute consistency or precision of a score and permits the calculation of confidence intervals (CIs). It is interpreted using the same units of measurement as used in the ACL-QOL score.[26,27] The 32 items in the ACL-QOL are scored using a VAS and converted to an overall score out of 100. In other words, each item in the ACL-QOL is intended to possess a weighting worth 1/32 and the final score are equivalent to a percentage score out of 100. As a result, the SEM should be interpreted on the scale out of 100.

Responsiveness Testing

A 1-way analysis of variance (ANOVA) was conducted to test the responsiveness of the ACL-QOL at 4 time periods (initial consultation, 6, 12, and 24 months postoperatively). A distribution-based model of responsiveness was calculated through effect size. Effect size was calculated manually by the Eta squared formula: sum of squares between groups/total sum of squares.

A convenience sample (n = 40) of the larger sample (n = 579) of patients completed a 7-point global rating scale and the ACL-QOL at minimum 24 months postoperatively. The convenience sample was selected by asking all patients who completed 24-month follow-up to complete the global rating scale in addition to the ACL-QOL, and the first 40 completed responses were used for correlation. Patients were asked to select 1 of 7 categories indicating their postoperative status: 7, significantly better; 6, much better; 5, somewhat better; 4, no change; 3, somewhat worse; 2, much worse; 1, significantly worse (adopted from Greco et al[27]). A Pearson r correlation coefficient was used to determine the relationship between the ACL-QOL change score (before and after surgery) with the 7-point scale response acting as the criterion of comparison.[27] All data analysis was computed using SPSS 22.0 (Chicago).

Retrospective Analysis of the Anterior Cruciate Ligament–Quality of Life Literature Using the Consensus-Based Standards for the Selection of Health Status Measurement Instruments Guidelines

To better understand the psychometric and clinimetric soundness of the ACL-QOL, the literature was reviewed for articles up to July, 2014 that cited original study of Mohtadi.[1] Specifically, the Google Scholar index was used to identify articles that cited and more importantly used the ACL-QOL.[28] Google Scholar index was chosen as the tool for this aspect of the study because it has the citation tracking ability for individual articles. All English articles were reviewed manually to determine whether they addressed any of the COSMIN taxonomy items: (1) internal consistency, (2) reliability, (3) standard error of measurement or limits of agreement, (4) content validity, (5) structural validity, (6) hypothesis testing, (7) cross-cultural validation, (8) criterion validity, and (9) responsiveness. Articles that added to the 9 COSMIN taxonomy items were cataloged accordingly.

RESULTS

Patient Population

The patient population was 52.1% men. The average age was 27.81 years (SD: 6.9, range: 11.2-56.8). The average body mass index was 25.02 (SD: 3.1, range: 15-38). In 51.7% of cases, the affected limb side was the right. At the time of data collection for this study, questionnaires completed totaled preoperative (n = 579), 6-month postoperative (n = 446), 12-month postoperative (n = 280), and 24-month postoperative (n = 100). These totals reflect the fact that patients were in various stages of follow-up in their postoperative course and data collection was ongoing.

Descriptive data for the cohort are listed in Table 1. These data were used as a comparison of groups under the distribution-based responsiveness measures.

Content Validity

The Cronbach's alpha for the preoperative, 6, 12, and 24 months postoperative periods is listed in Table 2. A number of patients did not answer all items in the ACL-QOL and therefore, analysis was completed on the sample sizes specified in Table 2.

Floor and ceiling effect calculations are presented in Table 3. All ACL-QOL responses were used in the calculation of the floor and ceiling effects. Floor/ceiling effects used a percent score based on the questions answered.

Reliability

Test–retest reliability was performed on 31 patients who completed the ACL-QOL twice, less than 2 weeks apart. The ICC (2,k) for this sample was 0.60. The SEM for the ACL-QOL in this sample was 6.16. A CI calculated from the SEM is a measure of the minimal detectable change based on the 95% CI. In this study, the CI was 12.1. When applied to the mean ACL-QOL scores at each serial testing stage, the CIs ranged from 14.6 to 38.8/100 at the preoperative stage, 48.3 to 72.5/100 at 6 months, 58.5 to 82.7/100 at 12 months, and 65.4 to 89.6/100 at 24 months.

TABLE 1. Descriptive Data at Preoperative, 6-, 12-, and 24-Months Postoperative

Time Period	Sample Size	Mean	SD	Confidence Interval (95%)
Preoperative	579	26.7	13.7	26.6-27.8
6-month postoperative	446	60.4	17.2	58.8-62.0
12-month postoperative	280	70.6	17.6	68.6-72.7
24-month postoperative	100	77.5	17.7	74.0-81.0

TABLE 2. Cronbach's Alpha on the ACL-QOL for the Various Time Periods

Measurement Period	Cronbach's Alpha	No. Subjects Excluded Due to Missing Items	Final Sample Size Analyzed
Preoperative	0.93	93	486
6-month postoperative	0.95	185	261
12-month postoperative	0.96	58	222
24-month postoperative	0.98	18	82

Responsiveness

There was a statistically significant difference at the $P < 0.05$ level for ACL-QOL scores for the 4 groups: F (3, 1401) = 729.1, $P = 0.001$. Based on the Eta squared calculation, the effect size was 0.61, which would be considered a large effect.[29] Post hoc comparisons using the Tukey honestly significant difference test indicated that the mean score for the preoperative testing period (M = 26.72, SD = 13.71, CI, 25.60-27.84) was significantly different from the 6-month postoperative testing time (M = 60.41, SD = 17.16, CI, 58.81-62.00), the 12-month postoperative testing time (M = 70.64, SD = 17.55, CI, 68.57-72.70), and the 24-month postoperative testing time (M = 77.48, SD = 17.71, CI, 73.97-81.00). In addition, there were significant differences between the 6-month postoperative group and the 12-month and 24-month postoperative groups. Finally, there were significant differences between the 12- and 24-month postoperative groups. Based on their ACL-QOL scores, patients continued to make significant improvements over a 2-year period.

The anchor-based method of evaluating responsiveness used a convenience sample of 40 patients who took the ACL-QOL preoperatively and a mean of 27 months postoperatively (SD = 3.3 months). Patients rated their improvement of significantly better (n = 8), much better (n = 24), somewhat better (n = 12), or no better (n = 1). A Pearson r correlation coefficient compared the mean change score in the ACL-QOL between the preoperative and 24-month postoperative scores with the 7-point rating scale. There was a significant correlation between these 2 measures (0.47, $P < 0.05$, 95% CI, 0.07-0.87) indicating the 2 scores are measuring similar constructs of improvement. The CI is indicative of the smaller sample used for this measurement.

Retrospective Analysis of the Anterior Cruciate Ligament–Quality of Life Literature Using the Consensus-Based Standards for the Selection of Health Status Measurement Instruments Guidelines

From the time of its publication in 1998 until July 2014, the Mohtadi ACL-QOL measure was cited in 139 journal articles.[28] Of these studies, 17 added to the validity, reliability, and/or responsiveness to the ACL-QOL. A summary of these articles is listed in Table 4 along with the COSMIN checklist items that were completed in this study and those that are under consideration for future studies.

DISCUSSIONS AND CONCLUSIONS

The ACL-QOL is a disease-specific, patient-reported QOL measure that has been validated in both the ACL deficient and the ACL reconstructed knee, across a number of studies. This study added further psychometric and clinimetric analyses of the ACL-QOL in the form of content validity, reliability, and responsiveness. In 2015, use of the ACL-QOL as a PROM marks a 20-year milestone of development. Outcome measures must continually evolve, progress, and be tested on new populations because validity is often sample-specific.[34,35] Although further research needs to be completed on the ACL-QOL, the mounting evidence moves closer to enabling statements that this PROM is "valid" and "reliable," without qualification.

Many outcome measures used in clinical studies make emphatic conclusions that their tool "is valid," "is reliable," or "is responsive." It is an error to assume that an instrument "is" any of those things because validity is considered an interaction between the instrument and the environment or patients being measured.[35–37] Validity is an iterative process whereby tools cannot claim to be valid until they have had a number of studies across a number of populations and subjects, after which they can begin to make claims of "validity."[36,37] Over 20 years, there have been a number of studies that used the ACL-QOL and each of these has contributed to building its validity, reliability, and/or responsiveness. Only in retrospection, with an accumulation of studies, across different populations with various interventions, one can begin to make bolder, concrete claims about the validity of an instrument.[38]

TABLE 3. Floor and Ceiling Effects by Percentage and Time Period

Floor or Ceiling Effects	Preoperative Period (% of total n = 579)	6-Month Postoperative (% of total n = 446)	12-Month Postoperative (% of total n = 280)	24-Month Postoperative (% of total n = 100)
Top 15%	0	6.7	23.2	42
Top 20%	0	13.2	32.1	51
Top 25%	0	22.9	47.5	64
Top 30%	0.17	32.3	55.7	71
Bottom 15%	24.5	0.44	0	1
Bottom 20%	34.3	0.90	0.36	1
Bottom 25%	53.5	2.0	1.4	1
Bottom 30%	66.0	5.4	3.2	1

TABLE 4. Retrospective Analysis of the ACL-QOL Literature Using the COSMIN Guidelines

COSMIN Checklist	Statistical Analysis	Studies
Internal consistency	Cronbach's alpha	Present study
Reliability	*t* test	Present study
	ICC	test–retest measured by *t* test in Mohtadi, 1998[1]
		Bryant et al, 2006 (ICC 0.86 plus SEM)[50]
SEM and/or limits of agreement (LOA)	SEM	Present study
	LOA	Bryant et al, 2006[50]
Content validity	Descriptive consensus	Mohtadi, 1998[1]
	Floor and ceiling	Tanner et al, 2007[4]
	Frequency-importance product	
Structural validity	Confirmatory factor analysis	Future research
	Exploratory factor analysis	
Hypothesis testing	*t* test or ANOVA, effect size, anchor-based MCID	Naylor et al, 2013[21]
		Bryant et al, 2006[50]
		Marks et al, 2008[49]
		Hiemstra et al, 2007[47]
		Arneja et al, 2004[46]
		Birmingham et al, 2002[51]
		Meridick et al, 2008[53]
		Birmingham et al, 2008[52]
Cross cultural validity	Language-specific consensus methods	Lysholm and Tegner, 2007[6]
		Kvist et al, 2013[48]
Criterion validity	Pearson *r* correlation to "objective" measures	Hiemstra et al, 2007[47]
	ROC curve	Future research
Responsiveness	Distribution-based	Present study
	Effect size	
	Standardized response mean	
	Anchor-based	
	Global rating scale of improvement correlated to QOL measure	

The content validity data in this study continue to support the overall validity of the ACL-QOL. The Cronbach's alpha ranged from 0.93 to 0.98 indicating that the ACL-QOL possesses unidimensionality, but these high numbers also indicate that there is some overlap and potential redundancy of the items in the questionnaire. A factor analysis and item reduction would help to clarify this aspect of the ACL-QOL validity. Future research on item reduction should proceed cautiously although the content of the ACL-QOL has been shown to connect so well with patients.[4]

Traditional definitions of floor and ceiling effects are intended to determine whether an evaluation instrument has the capability to measure patients at both ends of the spectrum. The exact percentage of patients that is considered within normal floor and ceiling limits varies considerably in the literature from 15% to 33%.[30,39] In this study, there is some evidence of floor or ceiling effects at the various time periods. However, it is also obvious that patients had substantially lower ACL-QOL scores preoperatively and that their health status or QOL improved significantly after ACL reconstruction.

In the original publication of the ACL-QOL, test–retest reliability was assessed using a paired *t* test. However, common practice since the original study has been to calculate reliability with an ICC. The reliability analysis in the test–retest sample of this study demonstrated very moderate reliability with an ICC (2,k) of 0.60. A previous study from Bryant et al[40] had a much higher ICC (0.86), but the methods used were different than this study, making it challenging to compare the 2. Bryant et al[40] calculated the ACL-QOL score by averaging the mean score of the 5 domains in the ACL-QOL rather than weighting each item equally to provide a score out of 100. Furthermore, the purpose of the Bryant et al[40] study was to determine patients' ability to recall their health status retrospectively. The sample for the test–retest calculation in this study was acceptable but small, which may also account for the variation.

The SEM is considered a measure of stability over time, which means one can confidently generalize these results to new populations.[26,32,41] The SEM in this study was 6.16 and when compared to the Bryant et al[40] study (6.12) seems to be stable and consistent across 2 different studies and populations. The CI in this study (12.1) was calculated from the ICC and SEM, and thus, clinicians and researchers could be 95% confident the score measured from the patient is either plus or minus 12.1 points. These CIs are useful to clinicians, particularly when they need to interpret minimally important difference (MID) and minimally clinically important difference (MCID). SEM can be classified as a measure of MID and it fits under the umbrella of a distribution-based approach.[2,41] This study also used an ANOVA and subsequently effect size as another distribution-based approach to MID.

Terwee et al[42] identified 31 different types of responsiveness. The responsiveness measurements in this study were both distribution-based and anchor-based approaches. In addition to the SEM, the primary distribution-based approach used in this study was effect size, which would be classified as moderately strong (0.61). Another crude mechanism to determine MID is to take half of the SD of the scores.[25,35] In this study, the SD ranged from 13.7 to 17.7 and thus using half of an SD approach would result in a MID ranging from ~6.9 to ~8.9. When these figures are compared to the SEM measures, once again, the MID seems to be stable with approximately 6 points on the ACL-QOL measurement.

It is important to take a multimeasured approach to MCID.[2,33,43] The results of anchor-based method in this study demonstrate that those patients whose ACL-QOL scores improved also believed they had improved. Revicki et al[44] and Sloan et al[45] advocated for a minimum Pearson r correlation coefficient of 0.3 up to 0.5, respectively. This study certainly falls within that range. Of the 40 patients who completed the 7-point anchor-based comparison to their ACL-QOL score, only 1 patient or 2.5% scored their experience as less than an optimal outcome (ie, some level of improvement). Therefore, it is possible that the convenience sample available for this analysis did not provide a representative sample across the entire spectrum of the responses to treatment. Future research should aim to study a larger population with the anchor-based approach to responsiveness and thus would facilitate receiver operating characteristic (ROC) curves. Specifically, a large enough sample of patients who did not improve would help to determine a cutpoint for responsiveness and MCID for the ACL-QOL. In summary, this study has demonstrated that the ACL-QOL is highly responsive to change for patients both from an MID and MCID perspective.

Limitations of this study include the use of a convenience sample for both the distribution-based and anchor-based methods used to measure the responsiveness of the ACL-QOL. There were also a large number of patients who had not yet reached the 12- or 24-month follow-up time interval and therefore these ACL-QOL scores were not available for inclusion. Despite the lack of complete ACL-QOL scores for each timepoint, the statistical analyses demonstrate that the sample size was sufficient for all analyses that were completed.

This is the largest cohort to date used to assess the validity and reliability of the ACL-QOL. The ACL-QOL has now completed 8 of 9 criteria outlined in the COSMIN guidelines (Table 4). Only structural validity and part of criterion validity remain. In future research, it would be appropriate to factor analyze the ACL-QOL, particularly in light of the high Cronbach's alpha scores. The original intent of the Mohtadi ACL-QOL was that 5 domains or underlying constructs were being measured as an indicator of the patient's overall health and QOL. Once these constructs have been confirmed, further research may be warranted to adjust the items as not all questions are answered because some patients do not work and/or others do not compete in contact sports. In addition, reducing the number of items in the questionnaire, and therefore the time it takes patients to complete it, would facilitate further use of the ACL-QOL in both clinical and research settings.

Regarding the criterion validation testing outlined in Table 4, there has been some criterion testing completed where there was a positive correlation between the ACL-QOL and a more "objective" measure of flexor and extensor strength after reconstructive surgery.[47] However, some researchers advocate for the criterion validation to take place with the subjective or patient-guided opinions of treatment success such as the ROC curve.[3] Another important metric outlined in the COSMIN guidelines is cross-cultural validation or translation to new languages.[6,48] The ACL-QOL has been successfully converted into some languages, but conversion to additional new languages is important in the future.[27,28] In summary of the retrospective analysis inventory of COSMIN guidelines, if the ROC curve and factor analysis are completed in the future, all 9 of the metrics identified in the COSMIN guidelines would have been met. Perhaps then, more global, concrete statements of the ACL-QOL can be made such as "it is valid, reliable, and responsive."

Despite the number of studies that have "hypothesis tested" (1 of the 9 COSMIN guidelines), there are still many questions that remain unanswered for clinicians who are looking to assess outcomes for ACL treatments.[21,40,47,49–53] Should future ACL treatments focus on PROMs, more objective measures or some combination of both? Should the assessment tools be disease-specific or is a general tool such as the International Knee Documentation Committee sufficient? Should all aspects of QOL be measured or merely certain aspects? Should certain research questions be matched with specific outcome measures? Future hypothesis testing needs to frame research questions, define outcomes explicitly, and use outcome measurement tools that correspond with those questions being posed. Then, when higher levels of evidence are considered for ACL treatments, perhaps more than just pivot shift and KT-1000 arthrometer tests can be used to define success.[53]

In conclusion, this study has added to the validity, reliability, and responsiveness of the ACL-QOL. In light of the COSMIN guidelines, the ACL-QOL demonstrated many of the metrics important for high-quality instruments. Furthermore, the ACL-QOL is the only disease-specific, patient-reported, QOL outcome measure that is intended for ACL-injured knees. Other outcome measures certainly meet some of those metrics, but not necessarily all. Future research is required to continue the iterative development of the ACL-QOL.

REFERENCES

1. Mohtadi N. Development and validation of the quality of life outcome measure (questionnaire) for chronic anterior cruciate ligament deficiency. *Am J Sports Med.* 1998;26:350–359.
2. Wyrwich K, Norquist J, Lenderking W, et al, Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res.* 2013;22:475–483.
3. Garratt AM, Brealey S, Gillespie WJ; DAMASK Trial Team. Patient-assessed health instruments for the knee: a structured review. *Rheumatology (Oxford).* 2004;43:1414–1423.
4. Tanner SM, Dainty KN, Marx RG, et al. Knee-specific quality-of-life instruments: which ones measure symptoms and disabilities most important to patients? *Am J Sports Med.* 2007;35:1450–1458.
5. Marx RG. Knee rating scales. *Arthroscopy*. 2003;19:1103–1108.
6. Lysholm J, Tegner Y. Knee injury rating scales. *Acta Orthop.* 2007;78: 445–453.

7. Irrgang JJ, Anderson AF, Boland AL, et al. Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med.* 2001;29:600–613.
8. Irrgang JJ, Anderson AF. Development and validation of health-related quality of life measures for the knee. *Clin Orthop.* 2002;402:95–109.
9. Irrgang JJ, Anderson AF, Boland AL, et al. Responsiveness of the international knee documentation committee subjective knee form. *Am J Sports Med.* 2006;34:1567–1573.
10. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum.* 2001;45:384–391.
11. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28:88.
12. Shapiro ET, Richmond JC, Rockett SE, et al. The use of a generic, patient-based health assessment (SF-36) for evaluation of patients with anterior cruciate ligament injuries. *Am J Sports Med.* 1996;24:196–200.
13. Marx RG, Jones EC, Allen AA, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am.* 2001;83:1459–1469.
14. Briggs KK, Lysholm J, Tegner Y, et al. The reliability, validity, and responsiveness of the lysholm score and tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. *Am J Sports Med.* 2009;37:890–897.
15. Gustavsson A, Neeter C, Thomeé P, et al. A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc.* 2006;14:778–788.
16. Barber-Westin SD, Noyes FR, McCloskey JW. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees. *Am J Sports Med.* 1999;27:402–416.
17. Bombardier C, Melfi CA, Paul J, et al. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care.* 1995;33:AS131–AS144.
18. Heyworth B, Brophy R, Marx R. Scoring the knee. In: *The Knee Joint*. Springer-Verlag, Paris, France; 2012:95–105.
19. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med.* 1993;118:622–629.
20. Schipper H. Guidelines and caveats for quality of life measurement in clinical practice and research. *Oncology (Williston Park).* 1990;4:51–57; discussion 70.
21. Naylor AJ, Mohtadi NG, Chan DS, et al. Anterior cruciate ligament preservation during reconstructive surgery: does the extra surgical effort improve patient outcomes at one year? *Sports Orthop Traumatol.* 2013; 29:29–39.
22. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737–745.
23. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Med Res Methodol.* 2010;10:82.
24. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Med Res Methodol.* 2011;11:152; author reply 152.
25. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.
26. Weir J. Quantifying test-rest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19:231–240.
27. Greco NJ, Anderson AF, Mann BJ, et al. Responsiveness of the international knee documentation committee subjective knee form in comparison to the Western Ontario and McMaster Universities osteoarthritis Index, modified Cincinnati knee rating system, and short form 36 in patients with focal articular cartilage defects. *Am J Sports Med.* 2010;38:891–902.
28. Google Scholar. *Google Scholar Citation Index Methods*. Web site. https://scholar.google.ca/intl/en/scholar/citations.html#citations. Accessed June 23, 2013.
29. Cohen J. A power primer. *Psychol Bull.* 1992;112:155.
30. Chaudhuri KR, Martinez-Martin P, Brown RG, et al. The metric properties of a novel non-motor symptoms scale for Parkinson's disease: results from an international pilot study. *Mov Disord.* 2007;22:1901–1911.
31. Cella D, Bullinger M, Scott C, et al; Clinical Significance Consensus Meeting Group. Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life. *Mayo Clin Proc.* 2002;77:384–392.
32. Palta M, Chen HY, Kaplan RM, et al. Standard error of measurement of 5 health utility indexes across the range of health for use in estimating reliability and responsiveness. *Med Decis Making.* 2011;31:260–269.
33. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Prof.* 2005;28:172–191.
34. Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. *J Adv Nurs.* 2014;70:1970–1979.
35. Norman G, Wyrwich KW, Patrick DL. The mathematical relationship among different forms of responsiveness coefficients. *Qual Life Res.* 2007;16:815–822.
36. Messick S. Test validity: a matter of consequence. *Soc Indic Res.* 1998; 45:35–44.
37. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Meas Issues Pract.* 1995;14:5.
38. Messick S. Validity of psychological assessment. *Am Psychol.* 1995;50:741.
39. Paxton EW, Fithian DC, Stone ML, et al. The reliability and validity of knee-specific and general health instruments in assessing acute patellar dislocation outcomes. *Am J Sports Med.* 2003;31:487–492.
40. Bryant D, Norman G, Stratford P, et al. Patients undergoing knee surgery provided accurate ratings of preoperative quality of life and function 2 weeks after surgery. *J Clin Epidemiol.* 2006;59:984–993.
41. de Vet HC, Terluin B, Knol DL, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. *J Clin Epidemiol.* 2010;63:37–45.
42. Terwee C, Dekker F, Wiersinga W, et al. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res.* 2003;12:349–362.
43. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–415.
44. Revicki DA, Cella D, Hays RD, et al. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes.* 2006;4:70.
45. Sloan JA, Frost MH, Berzon R, et al. The clinical significance of quality of life assessments in oncology: a summary for clinicians. *Support Care Cancer.* 2006;14:988–998.
46. Arneja S, Froese W, MacDonald P. Augmentation of femoral fixation in hamstring anterior cruciate ligament reconstruction with a bioabsorbable bead: a prospective single-blind randomized clinical trial. *Am J Sports Med.* 2004;32:159–163.
47. Hiemstra LA, Webber S, MacDonald PB, et al. Contralateral limb strength deficits after anterior cruciate ligament reconstruction using a hamstring tendon graft. *Clin Biomech.* 2007;22:543–550.
48. Kvist J, Österberg A, Gauffin H, et al. Translation and measurement properties of the Swedish version of ACL-return to sports after injury questionnaire. *Scand J Med Sci Sports.* 2013;23:568–575.
49. Marks P, O'Donnell S, Yee G. A pilot clinical evaluation comparing the mitek bone–tendon–bone cross pin and bioabsorbable screw in anterior cruciate ligament reconstruction fixation, a randomized double blind controlled trial. *Knee*. 2008;15:168–173.
50. Bryant D, Dill J, Litchfield R, et al. Effectiveness of bioabsorbable arrows compared with inside-out suturing for vertical, reparable meniscal lesions: a randomized clinical trial. *Am J Sports Med.* 2007;35:889–896.
51. Birmingham TB, Kramer JF, Kirkley A. Effect of a functional knee brace on knee flexion and extension strength after anterior cruciate ligament reconstruction. *Arch Phys Med Rehabil.* 2002;83:1472–1475.
52. Birmingham TB, Bryant DM, Giffin JR, et al. A randomized controlled trial comparing the effectiveness of functional knee brace and neoprene sleeve use after anterior cruciate ligament reconstruction. *Am J Sports Med.* 2008;36:648–655.
53. Meredick RB, Vance KJ, Appleby D, et al. Outcome of single-bundle versus double-bundle reconstruction of the anterior cruciate ligament: a meta-analysis. *Am J Sports Med.* 2008;36:1414–1421.