Development of a Valid and Reliable Knee Articular Cartilage Condition–Specific Study Methodological Quality Score

Joshua D. Harris,*†‡ MD, Brandon J. Erickson,‡ MD, Gregory L. Cvetanovich,‡ MD, Geoffrey D. Abrams,‡§ MD, Frank M. McCormick,‡‖ MD, Anil K. Gupta,‡¶ MD, MBA, Nikhil N. Verma,‡ MD, Bernard R. Bach Jr,‡ MD, and Brian J. Cole,‡ MD, MBA

Investigation performed at the Rush University Medical Center, Chicago, Illinois, USA

Background: Condition-specific questionnaires are important components in evaluation of outcomes of surgical interventions. No condition-specific study methodological quality questionnaire exists for evaluation of outcomes of articular cartilage surgery in the knee.

Purpose: To develop a reliable and valid knee articular cartilage–specific study methodological quality questionnaire.

Study Design: Cross-sectional study.

Methods: A stepwise, a priori–designed framework was created for development of a novel questionnaire. Relevant items to the topic were identified and extracted from a recent systematic review of 194 investigations of knee articular cartilage surgery. In addition, relevant items from existing generic study methodological quality questionnaires were identified. Items for a preliminary questionnaire were generated. Redundant and irrelevant items were eliminated, and acceptable items modified. The instrument was pretested and items weighed. The instrument, the MARK score (Methodological quality of ARticular cartilage studies of the Knee), was tested for validity (criterion validity) and reliability (inter- and intraobserver).

Results: A 19-item, 3-domain MARK score was developed. The 100-point scale score demonstrated face validity (focus group of 8 orthopaedic surgeons) and criterion validity (strong correlation to Cochrane Quality Assessment score and Modified Coleman Methodology Score). Interobserver reliability for the overall score was good (intraclass correlation coefficient [ICC], 0.842), and for all individual items of the MARK score, acceptable to perfect (ICC, 0.70-1.000). Intraobserver reliability ICC assessed over a 3-week interval was strong for 2 reviewers ($\geq$0.90).

Conclusion: The MARK score is a valid and reliable knee articular cartilage condition–specific study methodological quality instrument.

Clinical Relevance: This condition-specific questionnaire may be used to evaluate the quality of studies reporting outcomes of articular cartilage surgery in the knee.

Keywords: knee; articular cartilage; methodological quality; level of evidence; questionnaire

Both patient-reported and clinician-measured outcome questionnaires are used to evaluate a patient's subjective impression and the clinician's objective assessment of the success of an intervention. Many different types of questionnaires to evaluate success exist: general health (eg, Short Form–36 [SF-36]),[36] joint-specific (eg, American Shoulder and Elbow Surgeons [ASES] form),[42] limb-specific (eg, Disabilities of the Arm, Shoulder, and Hand [DASH] score),[24] and disease- or condition-specific tools. Disease-specific questionnaires are the optimal instruments to measure the response to an intervention in specific conditions (eg, Western Ontario Shoulder Instability [WOSI] score for patients with shoulder instability or Knee Numeric-Entity Evaluation Score [KNEES-ACL] for patients with anterior cruciate ligament [ACL] deficiency).[7,27] In patients with articular cartilage disease of the knee, validated and reliable patient-reported outcomes are increasingly used to guide treatment recommendations. Although there is no consensus agreement on the gold standard outcome instrument in this patient cohort, the International Knee Documentation Committee (IKDC) subjective score,[16] Knee injury and Osteoarthritis Outcome Score (KOOS) subscores,[3,9] and Lysholm knee scores[29] have acceptable properties to be used in patients with knee articular cartilage disorders.

Just as patient-specific outcomes are quantitatively evaluated, so is the methodological quality of individual studies. Unfortunately, the quality of articular cartilage literature has been limited by several methodological deficiencies, mostly in study design.[25,45] More recent literature does, however, demonstrate significant improvements in study quality.[19] In the current era of patient satisfaction–driven outcome metrics used to rate cost-efficient physician performance and reimbursement, it is necessary to practice not only "evidence-based medicine" but also "high-quality evidence-based medicine."[22] Just as clinical outcomes are primarily judged by patient-reported, condition-specific

The Orthopaedic Journal of Sports Medicine, 2(2), 2325967113512606
DOI: 10.1177/2325967113512606

TABLE 1
Common Generic Methodological Quality Instruments Used in Orthopaedic Surgery[a]

Coleman Methodology Score
Quality Appraisal Tool
CONSORT
Jadad
CLEAR-NPT
Delphi List
Cochrane BJMTG Score
Detsky Quality Assessment Scale
Modified Coleman Methodology Score

[a]CONSORT, Consolidated Standards of Reporting Trials; CLEAR-NPT, checklist to evaluate a report of a nonpharmacological trial; BJMTG, Bone, Joint and Muscle Trauma Group.

instruments, so should the assessment of study methodological quality via condition-specific questionnaires. The problem with generic questionnaires (eg, whole-body, limb-/joint-specific) is that they are not specific enough and ask irrelevant questions to the condition being investigated. The same lack of specificity to knee articular cartilage surgery exists for common contemporary methodological quality instruments, such as the Coleman Methodology Score. This tends to cluster quality scores around each level of evidence without significant variance between studies.[19] Although other investigations have modified the Coleman Methodology Score to evaluate the quality of autologous chondrocyte implantation (ACI) studies,[31] no knee articular cartilage study–specific quality questionnaire exists that is based on all methodological quality scores and all types of articular cartilage surgery. The purpose of this study was to develop a reliable and valid knee articular cartilage–specific study methodological quality questionnaire. The novel questionnaire would be used to evaluate the quality of studies reporting outcomes of articular cartilage surgery in the knee.

MATERIALS AND METHODS

The framework for development of a condition-specific methodological quality questionnaire for knee articular cartilage studies has not been established. In fact, this type of questionnaire has not been utilized for any condition in orthopaedic surgery, sports medicine, or related fields. The salient steps for development include (1) identification of relevant items from existing surgical outcome study quality questionnaires (Table 1); (2) identification of all patient-, knee-/limb-, defect-, and intervention-specific parameters relevant to patients undergoing articular cartilage surgery of the knee; (3) identification of all measures of outcome assessment; (4) item generation; (5) item reduction; (6) instrument pretesting; (7) item weighting; and (8) evaluation of instrument validity and reliability.

Item Identification, Item Generation, and Item Reduction

Instrument items are questions that may be dichotomous or polytomous. The items included in the final instrument may only exist if identified at the earliest stage of instrument development. Thus, this stage is the most important in creation of a novel questionnaire. The purpose of the new condition-specific instrument was to assess the quality of surgical knee articular cartilage investigations. The methodological quality of a study incorporates many factors and includes, but is not limited to, study design, conduct, reporting, analysis, interpretation, and external validity or generalizability of the findings. It has been defined as the ability that a study design will generate unbiased results and approach "the truth."[1]

A recent systematic review was used to evaluate the quality of 194 articular cartilage studies in the knee using 9 different study methodological quality questionnaires.[19] There were 124 total individual items within the 9 questionnaires. Duplicate items were removed, leaving 31 unique or acceptably similar items. Within all 194 studies analyzed, each study's primary, secondary, and any exploratory outcome measures (eg, purpose[s], hypothesis[es], result[s]) were analyzed with respect to all patient-, knee-, limb-, and defect-specific characteristics. For example, if a study's primary outcome was the 2-year Lysholm follow-up after microfracture, then this outcome score (Lysholm) was counted. Clinical outcome measures were extracted from all 194 investigations, and duplicate or acceptably similar outcome measures were removed, leaving 23 distinct measures. Similarly, duplicate or

*Address correspondence to Joshua D. Harris, MD, Houston Methodist Center for Orthopaedics & Sports Medicine, 6550 Fannin Street, Smith Tower, Suite 2500, Houston, TX 77030, USA (e-mail: joshuaharrismd@gmail.com).
†Houston Methodist Center for Orthopaedics & Sports Medicine, Houston, Texas, USA.
‡Rush University Medical Center, Chicago, Illinois, USA.
§Orthopedic Surgery, Stanford University and Veterans Administration–Palo Alto, Palo Alto, California, USA.
‖Holy Cross Hospital, Oakland Park, Florida, USA.
¶Florida Orthopaedic Institute, Tampa, Florida, USA.
One or more of the authors has declared the following potential conflict of interest or source of funding: B.J.C. receives royalties from Arthrex, DJO, and Elsevier; is on the speaker's bureau of Genzyme; is a paid consultant for Zimmer, Arthrex, Carticept, Biomimetic, Allosource, and DePuy; receives research support from Arthrex, Smith & Nephew, DJO, Zimmer, DePuy, and Johnson & Johnson; receives royalties from the publishers Lippincott, Elsevier, WB Saunders, and Smith & Nephew; and is on medical publications boards for the *Journal of Bone and Joint Surgery*, *American Journal of Sports Medicine*, *Cartilage*, *Journal of Shoulder and Elbow Surgery*, *American Journal of Orthopedics*, Elsevier, International Committee AANA, Educational Committee AANA, and the AAOS Board. B.R.B. receives royalties from SLACK Inc and institutional research support from Arthrex, Ossur, Linvatec, and Smith & Nephew. N.N.V. receives royalties from Smith & Nephew; is on the speaker's bureau for Arthrosurface; is a paid consultant for Smith & Nephew and Arthrex; owns stock in Omeros; receives research support from Arthrex, Smith & Nephew, Athletico, ConMed Linvatec, Miomed, Mitek, and Arthrosurface; receives royalties from the publishers Vindico Medical-Orthopedics Hyperguide and *Arthroscopy*; is on the medical publishing boards of the *Journal of Knee Surgery*, *Arthroscopy*, and SLACK Inc; and is a board member for the Arthroscopy Association Learning Center.

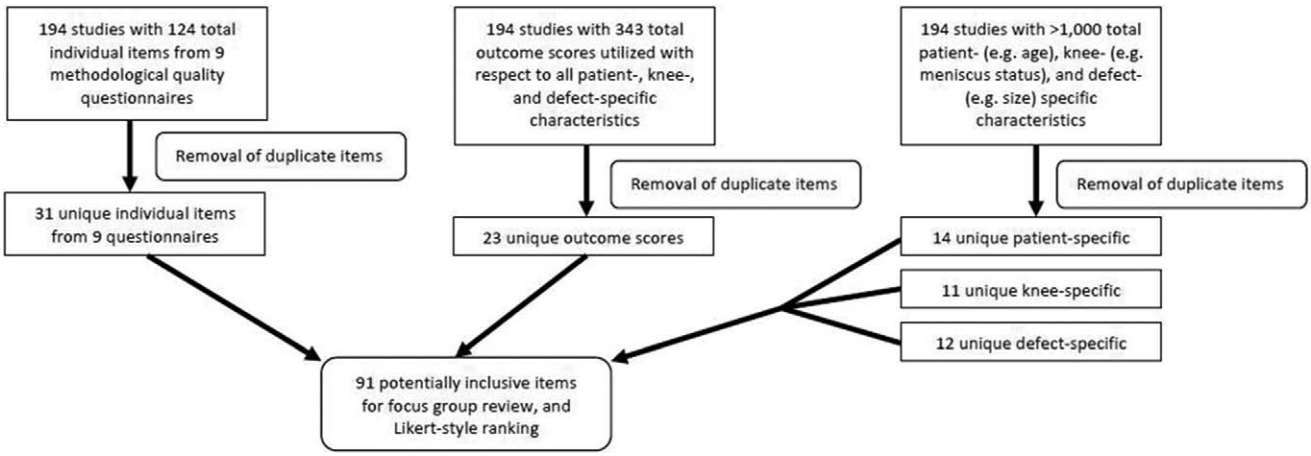

Figure 1. Flowchart illustrating the selection of 91 unique individual items via item identification, generation, and reduction for focus group review, analysis, and Likert-style ranking.

acceptably similar patient-, knee-/limb-, and defect-specific parameters were removed, leaving 14 (patient), 11 (knee/limb), and 12 (defect) unique items. The sum of the potentially inclusive items was 91 (Figure 1).

Focus Group Identification

To establish face validity, a focus group was created of 5 board-eligible sports medicine fellowship–trained orthopaedic surgeons and 3 board-certified sports medicine fellowship–trained (4, 9, and 24 years of experience) orthopaedic surgeons all familiar in treatment of articular cartilage injury in the knee. Face validity represents the ability of a questionnaire to measure what it is intended to measure using experts in the topic at hand.[8] This subjective property is considered valid if the items in the instrument represent all the relevant characteristics of the investigated condition (knee articular cartilage surgery) according to the subject matter experts.[39] All members of the focus group reviewed all potentially inclusive items (n = 91) and ranked them using a Likert-style questionnaire ranked 1 (strongly irrelevant) through 5 (strongly relevant) based on relevance of the item to the subject of articular cartilage surgery study methodological quality. Irrelevant and redundant items were removed. This procedure was repeated after a minimum of 1 week to avoid recall bias. Subtle modifications to individual items were done via group consensus. A total of 19 items were deemed relevant and included in the final questionnaire. Furthermore, to ensure optimal content validity, all focus group members were asked to provide any unmentioned items from the initial 91. No new or already present modified items were identified for new inclusion.

Pretesting

Pretesting the instrument is used to identify any weaknesses or problems with the final set of questions. All study authors were aware of the study purpose being development of a questionnaire that evaluates the quality of published articular cartilage studies and not the individual subjects' clinical outcomes. The instrument was administered to 2 study authors (1 board-eligible sports medicine fellowship–trained orthopaedic surgeon and 1 orthopaedic surgery resident physician) using 10 randomly selected studies from a group of 194 studies from a prior systematic review of methodological quality of articular cartilage studies.[19] Slight modifications were made to questions after consensus between authors after using the questionnaire for these 10 studies. The mean length of time needed to complete analysis of 1 study using the questionnaire (application of all 19 items to 1 study) was 4 minutes, 37 seconds.

Weighting

The authors chose to weight different items in the questionnaire differently, as some items contribute more to the overall importance of study quality. Although it was discussed to weight all items the same, this method was rejected via focus group decision as it would inevitably under- or overestimate the importance of some items. Based on the focus group's initial analysis of item relevance and importance, 19 items were selected for final inclusion in the instrument. Of these 19 items, the top 4 items were noted to be numerically the highest ranked importance (4.86, 4.57, 4.57, 4.57: length of follow-up, level of evidence, randomized trial, use of validated outcome scores, respectively). Furthermore, each focus group member was asked to also rank, in list order, their top 5 items for importance. The latter 4 were all included in all members' lists. Therefore, it was decided to weight these items with greater importance.

Validity and Reliability

A valid questionnaire is one that measures what it is supposed to measure. Validity may be established via 3 broad methods: content (face), criterion (concurrent), and construct validity. Face validity is established with an

expert-based focus group panel discussion and selection of appropriate items for inclusion in the instrument. Criterion validation is performed with comparison of the new instrument to a previously established gold standard. In the setting of articular cartilage literature, the new instrument, the MARK score (Methodological quality of ARticular cartilage studies of the Knee), is compared with other methodological quality instruments (Table 1). However, it must be recognized that no instrument has been designated as the "gold standard" methodological quality score. Correlation coefficients were used to examine the relationships between the described assessment tools. Values ranged from +1 (perfect positive correlation) to –1 (perfect negative correlation). Values greater than +0.7 were considered strongly positive correlations. A Shapiro-Wilk test indicated normal distribution for the MARK score and the Modified Coleman Methodology Score (MCMS). Comparisons were made using Pearson correlation coefficients for MARK versus MCMS, while all other MARK comparisons were made using Spearman correlation coefficients. Construct (convergent or discriminant) validation was not performed. Both inter- and intraobserver reliability was assessed via a 2-way mixed, single measures, intraclass correlation coefficient (ICC). ICC was acceptable if $\geq$0.70 (strong, $\geq$0.90; good, 0.80-0.89; fair, 0.70-0.79; poor, $\leq$0.69). Intraobserver reliability was assessed in 2 rounds with a 3-week interval to reduce recall bias via all 194 studies used in a recent articular cartilage systematic review.[19] Interobserver reliability was assessed for the overall MARK score and for each individual item via all 194 studies used in a recent articular cartilage systematic review.[19] Internal consistency of the data was assessed via Cronbach alpha. For all statistical analysis, $P < .05$ was statistically significant. SPSS version 18.0 (IBM, Armonk, New York, USA) was utilized for statistical analysis.

RESULTS

The MARK score has 19 items (Figure 2). The maximum raw score is 32; the minimum raw score is 0. The score is scaled to 100, with final score being: MARK score = (raw score/32) $\times$ 100. Three domains exist within the questionnaire: (1) study design methods (6 items), (2) subject or surgical demographics (5 items), and (3) outcome assessment (8 items).

Criterion Validity

The Pearson rank order correlation coefficient demonstrated a strongly positive correlation between MARK and MCMS ($r = 0.772$, $P < .001$). The Spearman rank order correlation coefficient demonstrated a strongly positive correlation between MARK and Cochrane Bone, Joint, and Muscle Trauma Group Quality Assessment Score ($r = 0.700$, $P = .002$), medium correlation between MARK and Delphi List ($r = 0.434$, $P = .081$), and small correlation between MARK and CLEAR-NPT (checklist to evaluate a report of a nonpharmacological trial) ($r = 0.128$, $P = .624$) and between MARK and Detsky ($r = 0.100$, $P = .702$).

Inter- and Intraobserver Reliability

Interobserver reliability ICC for all individual items of the MARK score were acceptable ($\geq$0.70) (Table 2). Interobserver reliability ICC for overall MARK score was acceptable. Intraobserver reliability ICC assessed over a 3-week interval was strong for both reviewers ($\geq$0.90).

DISCUSSION

A valid and reliable knee articular cartilage–specific study methodological quality questionnaire was developed. This questionnaire is not intended to evaluate patients' clinical outcomes but rather used to evaluate the quality of studies reporting outcomes of articular cartilage surgery in the knee. This instrument generates a raw score that is easily convertible to a MARK score (minimum, 0 points; maximum, 100 points). Given that patient-reported, condition-specific outcomes are among the best quantitative evaluations of the result of an intervention, so should the assessment of study methodological quality via condition-specific questionnaires. The MARK score is certainly not intended to replace but rather to supplement existing instruments and scores.

Each item on the MARK score intends to address relevant information that correlates with or is predictive of the outcome of surgical treatment of a chondral defect of the knee. Higher levels of evidence (item 1) should be more convincing to clinicians attempting to solve a clinical problem.[46] Although the best evidence study design is a high-quality randomized trial, sometimes this trial type is impossible (perhaps because a true control group does not exist, like in articular cartilage surgery) and "lower" levels of evidence are the best available. The presence of a financial conflict of interest (item 2), may present a bias to the reader that, even when disclosed, may affect outcome interpretation. However, industry funding is often necessary to support research, given constraints and limits imposed by public and government sources for distribution and utilization. Nonetheless, the level of evidence of industry-funded research is lower than that found in non-industry-funded research or investigations funded by government or public sources.[40] Presentations of authors with financial conflicts of interest have been found to more likely describe positive findings.[41] More recently, publications with conflicts of interest present have also contributed to the increase in negative outcomes of studies reported in the literature with later publication dates.[23] In a recent systematic review, there was a significant increase in the number of studies that adequately reported either the presence or absence of a financial conflict of interest, reflecting journal editors' and study authors' recognition of the impact that these conflicts may have on study outcomes.

The Jadad scale is a very simple (3 items) tool used to evaluate the quality of a randomized trial. Although sometimes criticized because of its simplicity, it is the most widely used quality tool internationally and was integrated into the MARK score (item 3). Sample size calculation (power analysis) is necessary to determine the minimum

1) Level of evidence
   a. I = 3
   b. II = 2
   c. III = 1
   d. IV = 0
2) Does study report presence of financial COI?
   a. Yes = 0
   b. No = 1
   c. COI information not reported = 0
3) Is study a randomized controlled trial (RCT)?
   a. Yes = 1
      i. If yes, was power analysis performed?
         1. Yes = 1
         2. No = 0
      ii. If yes, was one comparison group comprised of patients undergoing no treatment to chondral defect (i.e. control group)?
         1. Yes = 1
         2. No = 0
   b. No = 0
4) Was study sample size >30 subjects per group?
   a. Yes = 1
   b. No = 0
5) Was study enrollment rate, inclusion/exclusion criteria fully and transparently described?
   a. Yes = 1
   b. No = 0
6) Length of clinical follow-up
   a. < 2 years = 0
   b. 2-5 years = 2
   c. 5-10 years = 4
   d. >10 years = 6
7) Did subjects undergo any prior knee surgery?
   a. No = 2
   b. Prior surgeries not reported = 0
   c. Yes - Prior cartilage, realignment (coronal plane, patellofemoral), meniscal, or ligament surgery = 0*
   d. Yes - But prior surgery (includes debridement, ACI biopsy) includes none of the above = 1
8) Did subjects undergo concomitant surgery?
   a. No = 2
   b. Yes = 1 (meniscal transplant/repair, realignment osteotomy, ligament surgery)
   c. Yes = 0 (meniscectomy, untreated ligamentous or alignment abnormality)
9) Was surgical technique appropriately described?
   a. Yes = 1
   b. No = 0
10) Was post-operative rehabilitation appropriately described?
   a. Yes = 1
   b. No = 0
11) Were outcomes reported as separate groups for individual knee compartments (patellofemoral versus tibiofemoral) or assimilated as one group?
   a. Yes = 1
   b. No = 0
12) Were outcomes of patients with both chondral and osteochondral defects reported separately?
   a. Yes = 1
   b. No = 0
13) Were subjects and defects appropriately** described (age, gender, body mass index, smoking status, duration of symptoms, defect size, depth, location)?
   a. Yes = 1
      i. If yes, were the compared groups similar?***
         1. Yes = 1
         2. No = 0
   b. No = 0
14) Did study report general health score (e.g. SF-36) and cartilage-validated outcome score (IKDC, KOOS, Lysholm)?
   a. Both general health and cartilage score = 2
   b. Either general health or cartilage score = 1
   c. Neither = 0
15) Did study report patient satisfaction questionnaire (was patient satisfied / would they have same surgery again)?
   a. Yes = 1
   b. No = 0
16) Was an independent observer utilized for post-operative objective assessments (clinical, imaging, or histologic)?
   a. Yes = 1
   b. No = 0
17) Were post-operative complications and re-operations reported?
   a. Yes = 1
   b. No = 0
18) Were x-rays obtained post-operatively to determine presence / absence of findings of arthritis?
   a. Yes = 1
   b. No = 0
19) Was advanced cartilage-specific MRI (dGEMRIC, T1-rho, T2-mapping, sodium [Na], MOCART, WORMS) obtained pre- or post-operatively?
   a. Yes = 1
   b. No = 0

MARK Score = (Raw score/32)x100.

Raw score 0 – 32 (scaled score)
24-32 = Excellent (75-100)
16-23 = Good (50-75)
8-15 = Fair (25-50)
0-7 = Poor (<25)

*Includes: microfracture, drilling, abrasion, ACI, osteochondral autograft or allograft, cell-based surgical treatments, high tibial, distal femoral, or tibial tubercle osteotomies, meniscectomy, meniscal repair or transplant.
**4 or more items yields 1 point; 3 or fewer items yield 0 points.
***If one group of patients only (e.g. retrospective case series), then cannot answer "yes" to this question.

Figure 2. MARK score. COI, conflict of interest; ACI, autologous chondrocyte implantation; SF-36, Short Form–36; IKDC, International Knee Documentation Committee; KOOS, Knee injury and Osteoarthritis Outcome Score; dGEMRIC, delayed Gadolinium-Enhanced MRI of Cartilage; MOCART, Magnetic resonance Observation of CArtilage Repair Tissue; WORMS, Whole-ORgan MRI Score.

TABLE 2
MARK Score Measurements and Inter- and Intraobserver Reliability Using 194 Recent Surgical Studies of Articular Cartilage in the Knee[a]

	Round 1	Round 2
MARK score, mean ± SD		
Overall	36.9 ± 11.8	37.2 ± 11.5
Level 1	57.8 ± 12.3	57.8 ± 12.3
Level 2	46.6 ± 8.56	47.2 ± 8.91
Level 3	33.8 ± 10.0	33.8 ± 10.1
Level 4	34.4 ± 8.72	34.4 ± 8.75
Intraobserver reliability, ICC (95% CI)		
Reviewer 1	0.995 (0.986, 0.998); Cronbach α = 0.995	
Reviewer 2	0.990 (0.973, 0.996); Cronbach α = 0.995	
Interobserver reliability, ICC (95% CI)	0.842 (0.564, 0.943); Cronbach α = 0.845	
Interobserver reliability per individual item, ICC (95% CI)		
1	0.875 (0.655, 0.955)	
2	0.776 (0.382, 0.919)	
3	0.792 (0.426, 0.925)	
4	1.000 (1.000, 1.000)	
5	0.789 (0.419, 0.924)	
6	0.977 (0.937, 0.992)	
7	0.842 (0.563, 0.943)	
8	0.646 (0.021, 0.872)	
9	0.789 (0.419, 0.924)	
10	0.789 (0.419, 0.924)	
11	0.875 (0.655, 0.955)	
12	0.786 (0.410, 0.923)	
13	0.776 (0.382, 0.919)	
14	0.828 (0.525, 0.938)	
15	1.000 (1.000, 1.000)	
16	0.775 (0.378, 0.918)	
17	0.907 (0.743, 0.966)	
18	0.852 (0.592, 0.946)	
19	0.759 (0.333, 0.913)	

[a]CI, confidence interval; ICC, intraclass correlation coefficient; SD, standard deviation.

sample size needed so that an investigation is powered to detect an effect of a given size. Of the 16 randomized trials analyzed in this review, only 5 (31%) performed an appropriate power analysis based on validated cartilage outcome scores. The mean number of the minimum sample sizes reported in these 5 studies was 30, hence the selection of 30 subjects per group as the threshold found in item 4. Transparent subject enrollment with explicit inclusion and exclusion criteria is a necessary important component of all studies, including cartilage research (item 5). This was illustrated in that 8 of 9 study quality questionnaires used in a recent review used it as an individual item.[19] Unfortunately, it was also identified as a significant weakness in most articular cartilage studies due to seldom reporting.

Length of clinical follow-up is a key component of the assessment of outcome of articular cartilage surgery, as it reflects the durability of the intervention. In other words, can the surgery halt the progression of degenerative changes? Many generic quality questionnaires (eg, MCMS) report 2 years as "long-term" follow-up. However, for ACI, it takes 2 years for tissue maturation based on previous investigations utilizing second look arthroscopy and biopsy.[5,12] Furthermore, it can take athletes up to 2 years to recover following ACI.[17] It is also known that fibrocartilage wear and durability can lead to declines in clinical outcome beginning as early as 18 to 24 months after microfracture.[13,30,32,38] Thus, 2 years is clearly insufficient for defining "long-term" follow-up. Weighting of this item in the MARK score (item 6) reflects the importance of long-term follow-up in being able to adequately assess the outcome of the intervention performed. This is also evident in the weighting of the item during the weighting stage of instrument development. Prior (item 7)[10,34,37] and concomitant (item 8)[14,18,20] surgeries have the potential to significantly influence cartilage surgery. Since most of the articular cartilage literature permits these other procedures, the true outcome of the intervention performed is confounded.

Transparent study reporting is necessary for all clinical studies with satisfactory descriptions of surgical technique (item 9), postoperative rehabilitation (item 10), and subjects (item 13), as many of these factors influence outcomes.[2] Factors relevant to ACI include biopsy location, size, and number of cells; biopsy age at time of implantation; cover choice (periosteum, type I-III collagen membrane, 3-dimensional scaffold); defect shouldering; verticality of walls; and management of calcified cartilage zone (kept intact).[2,11,12,43] Regarding microfracture,

placement of holes peripheral versus central, defect shouldering, verticality of walls, and calcified cartilage zone removal are all relevant.[11,43,44] For both autograft and allograft osteochondral transfer, the number, size, depth, donor site, and placement (flush/proud/recessed) of plugs warrant mention.[15,21,35] For allograft, sterilization and storage methods are introduced.[6,33] Following cartilage surgery, rehabilitation (especially joint motion and weightbearing status) plays a vital role and merits discussion.[28,44] Patient description of age, sex, body mass index, smoking status, duration of symptoms, and prior treatments all influence cartilage surgery outcomes.[2,26] Given the distinct differences between anatomy, biomechanics, and surgical techniques of patellofemoral and tibiofemoral lesions, their outcomes should be reported separately (item 11). Similarly, chondral and osteochondral defects behave differently and should be reported as such (item 12).

The importance of using patient-reported outcomes with optimal psychometric properties (item 14) and assessing patient satisfaction (item 15) cannot be overemphasized, as they provide the best evaluation of the true outcome of the intervention. To reduce detection bias, use of an independent examiner is necessary in measuring clinical, radiographic, or histological outcomes (item 16). As with any surgical technique, a safety profile must be known and complications and reoperations reported (item 17), especially given the high rate of reoperation following certain cartilage techniques.[18,20] Although infrequently reported in short- and medium-term investigations, radiographic examination is necessary to monitor for the potential development and progression of osteoarthritis (item 18). Similarly, newer advanced magnetic resonance imaging techniques have the ability to detect both articular cartilage structure and physiology both before and after surgery and also predict clinical outcome (item 19).[4]

A recent study quality investigation exclusively evaluating ACI analyzed 18 studies and 731 subjects based on use of only the Coleman Methodology Score (high internal validity).[31] Thus, the latter is a highly specific questionnaire appropriate to analyze studies on ACI. The current questionnaire is, in no way, "better" or "worse" than the ACI-specific questionnaire. The basis for the current questionnaire was a systematic review of 194 studies and nearly 12,000 subjects on all articular cartilage surgery with production of a new questionnaire based on use of 9 separate methodological quality questionnaires (with a concurrent increase in generalizability and external validity).

Limitations

The MARK score was appropriately developed and tested for reliability and validity. Nonetheless, limitations in it and its use may exist. First and foremost, it must be emphasized that this instrument is intended to measure the quality of a study, not the outcomes of the study. Thus, it measures not just the design and performance of the study but also the actual reporting of the study. It does not measure what the patient reports or the clinician measures as the outcome of individual subjects. The measurement of study quality via the MARK score was all made from a separate study.[19] Measurement of the MARK score in this study and the other 9 instruments from the latter study are subject to performance and detection bias from individual reviewers. The latter study, although inclusive of 194 studies, only included studies from the past 10 years of articular cartilage research, thus potentially creating selection bias. Another potential source of selection bias is that the reviewers and focus group members were orthopaedic surgeons and not biostatisticians or epidemiologists. The lack of construct validity (convergent or discriminant) also represents a limitation. However, given the multifactorial nature of knee articular cartilage surgery, the authors utilized face validation using our cartilage surgeons' focus group with nearly 5 decades of experience for the purposes of this investigation. In addition to face validity, the MARK score demonstrated a strong correlation with the MCMS and the Cochrane Bone, Joint, and Trauma Muscle Group Quality Assessment Score (criterion validity). The Cochrane score is currently used by the largest evidence-based medicine network for the assessment of study quality in the world, the Cochrane Collaboration. The MCMS is currently one of the most popular assessment tools in orthopaedic surgery. However, only moderate or small correlations were demonstrated for the other quality assessment instruments used. It is possible that, despite efforts to appropriately weight each item on the questionnaire, weighting emphasis was misplaced on certain items to overemphasize (greater point value assignment) or underemphasize (smaller point value) their importance or contribution to the overall quality of the study. Furthermore, there may be slight redundancy in certain items on the questionnaire (eg, items 1 and 3, level of evidence and study randomization). However, this is a method that places greater emphasis on higher quality study design. Furthermore, question content overlap is a constituent of validation that attempts to avoid or minimize question content redundancy without loss of necessary content. It is certainly not feasible for authors and/or readers to assess a single or multiple studies via several different questionnaires. Therefore, the most appropriate type of questionnaire for each condition being studied should be used, with reasonable amounts of time required for reading the investigation and grading it using the instrument. The field of articular cartilage surgery in the knee is rapidly growing, as is the associated international financial investment. Thus, the best gauge of the quality of the literature, an articular cartilage–specific methodology score, is a useful and beneficial development. Future research may apply this same methodology to other common conditions.

CONCLUSION

We believe a valid and reliable knee articular cartilage condition–specific study methodological quality instrument, the MARK score, has been developed. This condition-specific questionnaire may be used to evaluate the quality of studies reporting outcomes of articular cartilage surgery in the knee.

REFERENCES

1. Arrive L, Renard R, Carrat F, et al. A scale of methodological quality for clinical studies of radiologic examinations. *Radiology*. 2000;217:69-74.
2. Behery OA, Harris JD, Karnes JM, Siston RA, Flanigan DC. Factors influencing the outcome of autologous chondrocyte implantation: a systematic review. *J Knee Surg*. 2013;26:203-211.
3. Bekkers J, deWindt T, Raijmakers N, Dhert W, Saris D. Validation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) for the treatment of focal cartilage lesions. *Osteoarthritis Cartilage*. 2009; 17:1434-1439.
4. Blackman AJ, Smith MV, Flanigan DC, Matava MJ, Wright RW, Brophy RH. Correlation between magnetic resonance imaging and clinical outcomes after cartilage repair surgery in the knee: a systematic review and meta-analysis. *Am J Sports Med*. 2013;41:1426-1434.
5. Brun P, Dickinson SC, Zavan B, Cortivo R, Hollander AP, Abatangelo G. Characteristics of repair tissue in second-look and third-look biopsies from patients treated with engineered cartilage: relationship to symptomatology and time after implantation. *Arthritis Res Ther*. 2008;10:R132.
6. Chahal J, Gross AE, Gross C, et al. Outcomes of osteochondral allograft transplantation in the knee. *Arthroscopy*. 2013;29:575-588.
7. Comins JD, Krogsgaard MR, Brodersen J. Development of the Knee Numeric-Entity Evaluation Score (KNEES-ACL): a condition-specific questionnaire. *Scand J Med Sci Sports*. 2013;23:e293-e301.
8. Comins JD, Krogsgaard MR, Brodersen J. Ensuring face validity in patient-related outcome scores—a matter of content. *Knee*. 2013; 20:72-78.
9. Engelhart L, Nelson L, Lewis S, et al. Validation of the Knee Injury and Osteoarthritis Outcome Score subscales for patients with articular cartilage lesions of the knee. *Am J Sports Med*. 2012;40:2264-2272.
10. Fauno P, Nielsen AB. Arthroscopic partial meniscectomy: a long-term follow-up. *Arthroscopy*. 1992;8:345-349.
11. Frisbie DD, Morisset S, Ho CP, Rodkey WG, Steadman JR, McIlwraith CW. Effects of calcified cartilage on healing of chondral defects treated with microfracture in horses. *Am J Sports Med*. 2006;34: 1824-1831.
12. Gikas P, Morris T, Carrington R, Skinner J, Bentley G, Briggs T. A correlation between the timing of biopsy after autologous chondrocyte implantation and the histological appearance. *J Bone Joint Surg Br*. 2009;91:1172-1177.
13. Gobbi A, Karnatzikos G, Kumar A. Long-term results after microfracture treatment for full-thickness knee chondral lesions in athletes [published online September 20, 2013]. *Knee Surg Sports Traumatol Arthrosc*. doi:10.1007/s00167-013-2676-8.
14. Gomoll AH, Kang RW, Chen AL, Cole BJ. Triad of cartilage restoration for unicompartmental arthritis treatment in young patients: meniscus allograft transplantation, cartilage repair and osteotomy. *J Knee Surg*. 2009;22:137-141.
15. Gudas R, Gudaite A, Pocius A, et al. Ten-year follow-up of a prospective, randomized clinical study of mosaic osteochondral autologous transplantation versus microfracture for the treatment of osteochondral defects in the knee joint of athletes. *Am J Sports Med*. 2012; 40:2499-2508.
16. Hambly K, Griva K. IKDC or KOOS? Which measures disability symptoms and disabilities most important to postoperative articular cartilage repair patients? *Am J Sports Med*. 2008;36:1695-1704.
17. Harris JD, Brophy RH, Siston RA, Flanigan DC. Treatment of chondral defects in the athlete's knee. *Arthroscopy*. 2010;26:841-852.
18. Harris JD, Cavo M, Brophy R, Siston R, Flanigan D. Biological knee reconstruction: a systematic review of combined meniscal allograft transplantation and cartilage repair or restoration. *Arthroscopy*. 2011;27:409-418.
19. Harris JD, Erickson BJ, Abrams GD, et al. Methodologic quality of knee articular cartilage studies. *Arthroscopy*. 2013;29:1243.e5-1252.e5.
20. Harris JD, Siston RA, Brophy RH, Lattermann C, Carey JL, Flanigan DC. Failures, re-operations, and complications after autologous chondrocyte implantation—a systematic review. *Osteoarthritis Cartilage*. 2011;19:779-791.
21. Harris JD, Solak KK, Siston RA, Litsky A, Richards J, Flanigan DC. Contact pressure comparison of proud osteochondral autograft plugs versus proud synthetic plugs. *Orthopedics*. 2011;34:97.
22. Hospital Care Quality Information from the Consumer Perspective. 2012. http://www.hcahpsonline.org. Accessed May 6, 2012.
23. Hsu JE, Liu SY, Lee GC. Can we trust studies published by authors with financial conflicts? Using the decline of metal-on-metal total hip arthroplasty to investigate. *J Arthroplasty*. 2012;27:41.e1-45.e1.
24. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med.* 1996;29:602-608.
25. Jakobsen RB, Engebretsen L, Slauterbeck JR. An analysis of the quality of cartilage repair studies. *J Bone Joint Surg Am*. 2005;87: 2232-2239.
26. Kanneganti P, Harris JD, Brophy RH, Carey JL, Lattermann C, Flanigan DC. The effect of smoking on ligament and cartilage surgery in the knee: a systematic review. *Am J Sports Med*. 2012;40:2872-2878.
27. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med*. 1998;26:764-772.
28. Knapik DM, Harris JD, Pangrazzi G, et al. The basic science of continuous passive motion in promoting knee health: a systematic review of studies in a rabbit model. *Arthroscopy*. 2013;29:1722-1731.
29. Kocher MS, Steadman JR, Briggs KK, Sterett WI, Hawkins RJ. Reliability, validity, and responsiveness of the Lysholm Knee Scale for various chondral disorders of the knee. *J Bone Joint Surg Am*. 2004;85A: 1139-1145.
30. Kon E, Filardo G, Berruto M, et al. Articular cartilage treatment in high-level male soccer players: a prospective comparative study of arthroscopic second-generation autologous chondrocyte implantation versus microfracture. *Am J Sports Med*. 2011;39:2549-2557.
31. Kon E, Verdonk P, Condello V, et al. Matrix-assisted autologous chondrocyte transplantation for the repair of cartilage defects of the knee: systematic clinical data review and study quality analysis. *Am J Sports Med*. 2009;37(suppl 1):156S-166S.
32. Krych AJ, Harnly HW, Rodeo SA, Williams RJ 3rd. Activity levels are higher after osteochondral autograft transfer mosaicplasty than after microfracture for articular cartilage defects of the knee: a retrospective comparative study. *J Bone Joint Surg Am*. 2012;94:971-978.
33. Krych AJ, Robertson CM, Williams RJ 3rd. Return to athletic activity after osteochondral allograft transplantation in the knee. *Am J Sports Med*. 2012;40:1053-1059.
34. Li RT, Lorenz S, Xu Y, Harner CD, Fu FH, Irrgang JJ. Predictors of radiographic knee osteoarthritis after anterior cruciate ligament reconstruction. *Am J Sports Med*. 2011;39:2595-2603.
35. Matricali GA, Dereymaeker GP, Luyten FP. Donor site morbidity after articular cartilage repair procedures: a review. *Acta Orthop Belg*. 2010;76:669-674.
36. McHorney C, Ware J, Raczek A. The MOS 36-item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247-263.
37. Minas T, Gomoll A, Rosenberger R, Royce R, Bryant T. Increased failure rate of autologous chondrocyte implantation after previous treatment with marrow stimulation techniques. *Am J Sports Med*. 2009;37:902-908.
38. Mithoefer K, Williams RJ 3rd, Warren RF, Wickiewicz TL, Marx RG. High-impact athletics after knee articular cartilage repair: a prospective evaluation of the microfracture technique. *Am J Sports Med*. 2006;34:1413-1418.
39. Mosier C. A critical examination of the concepts of face validity. *Educ Psychol Meas*. 1947;7:191-205.
40. Noordin S, Wright JG, Howard A. Relationship between declared funding support and level of evidence. *J Bone Joint Surg Am*. 2010; 92:1647-1651.

41. Okike K, Kocher MS, Mehlman CT, Bhandari M. Conflict of interest in orthopaedic research. An association between findings and funding in scientific presentations. *J Bone Joint Surg Am*. 2007;89:608-613.
42. Richards R, An K, Bigliani L, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg*. 1994;3:347-352.
43. Rudd RG, Visco DM, Kincaid SA, Cantwell HD. The effects of beveling the margins of articular cartilage defects in immature dogs. *Vet Surg*. 1987;16:378-383.
44. Steadman J, Rodkey W, Briggs K. Microfracture to treat full-thickness chondral defects: surgical technique, rehabilitation, and outcomes. *J Knee Surg*. 2002;15:170-176.
45. Worthen J, Waterman CP, Davidson PA, Lubowitz JH. Limitations and sources of bias in clinical knee cartilage research. *Arthroscopy*. 2012; 28:1315-1325.
46. Wright J, Swiontkowski M, Heckman J. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003;85:1-3.