

Interrater Agreement of an Arthroscopic Anterior Cruciate Ligament Tear Classification System

BEAR-MOON*†

Investigation performed at Cleveland Clinic, Cleveland, Ohio, USA

Background: Anterior cruciate ligament (ACL) rupture is the most common ligament injury treated surgically by orthopaedic surgeons. The gold standard for the treatment of the majority of primary ACL tears is ACL reconstruction. However, novel methods of repair, such as bridge-enhanced ACL repair (BEAR), are currently being investigated as alternatives to reconstruction. To assess patients for midsubstance repair suitability, clarify the prognostic implications of injury location and damage, and evaluate the results of a repair technique, it is important to have a baseline classification system or grading scale that is reproducible across surgeons, particularly for multicenter collaboration. Currently, no such system or scale exists.

Purpose: To develop an arthroscopic ACL tear classification system and to evaluate its interobserver reliability.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: Eleven fellowship-trained orthopaedic surgeon investigators reviewed 75 video clips containing arthroscopic evaluation of a torn ACL and then completed the 6-question ACL Pathology Evaluation Form. Agreement statistics including exact agreement, Fleiss κ , Gwet agreement coefficient 1 (AC1), and Gwet AC2 were then calculated to assess interobserver reliability.

Results: In aggregate, the multiple assessments of observer reproducibility revealed that surgeon participants in this study, when evaluating the same injury, agreed roughly 80% of the time on whether (1) at least 50% of the tibial footprint remained, (2) the remaining tibial stump was ≥ 10 mm, and (3) the injury was therefore reparable using the BEAR procedure. Participants also agreed roughly 60% of the time on exactly how many suturable bundles were available. These characteristics are believed to be most important, among those studied, in determining whether a torn ACL is amenable to midsubstance repair.

Conclusion: This study is the first of its kind to demonstrate the interobserver reliability of arthroscopic classification of ACL tears. We have demonstrated that this classification system, though not ideally reproducible, is reliable enough across surgeons at multiple institutions for use in multicenter studies.

Registration: NCT03776162 (ClinicalTrials.gov identifier).

Keywords: ACL tear; ACL repair; ACL reconstruction; midsubstance repair; interrater agreement; reliability

Acute anterior cruciate ligament (ACL) rupture is the most common ligament injury treated surgically by orthopaedic surgeons.^{7,20} Currently, the gold standard for the treatment of primary ACL tears is a reconstruction utilizing predominantly autograft patellar tendon or hamstring tendon, with allografts reserved for middle-aged patients. The choice of graft is based on patient-specific factors, which have also been incorporated into risk calculators and nomograms.^{8,11,19,22} However, novel methods of repair are currently being investigated as alternatives to reconstruction. The potential advantages of repair are avoidance of autograft harvest; reduced morbidity; and maintenance of a “normal” tibial footprint, which can have potentially better kinematics and preserve proprioceptive nerve supply.

Moreover, bridge-enhanced ACL repair (BEAR)—which combines suture repair of the ACL with placement of a specific extracellular matrix scaffold to facilitate ligament healing into the gap between the torn ACL ends^{14,15}—has reduced posttraumatic osteoarthritis as compared with reconstruction in translational large animal models.^{3,14,15,17,18} Other primary repair techniques, with or without “internal stabilization” and with or without biologic enhancement, are also being used clinically.^{1,10,13,16}

ACL ruptures vary in the location of the ligamentous disruption (eg, pure avulsion from the femoral insertion, proximal-third midsubstance rupture, distal-third midsubstance rupture), the proportion of the tibial footprint that remains intact, the damage to the remaining ACL stump, and the residual ability of the stump to hold a suture for repair. These factors affect whether a repair can be performed (ie, primary repair vs reconstruction vs BEAR) and may also influence repair healing.

The Orthopaedic Journal of Sports Medicine, 8(12), 2325967120966323
DOI: 10.1177/2325967120966323
© The Author(s) 2020

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

To assess patients for repair suitability, clarify the prognostic implications of injury location and damage, and evaluate the results of a repair technique, it is important to have a baseline classification system or grading scale that is reproducible across surgeons, particularly for multicenter collaborations. Such a system was developed for meniscal and articular cartilage for use in 2 prospective longitudinal cohorts funded by the National Institutes of Health (NIH): MARS (Multicenter ACL Revision Study) and MOON (Multicenter Orthopaedics Outcomes Network).^{4,12} To our knowledge, no such arthroscopic ACL tear classification system currently exists. In this study, we aimed to develop such a system and evaluate its interobserver reliability.

We hypothesized that fellowship-trained sports medicine orthopaedic surgeons who are coinvestigators in an NIH-funded randomized controlled trial on ACL repair would exhibit moderate to high agreement (reliability) in classifying ACL tears based on a series of predefined qualities: ACL tibial stump length, percentage of ACL tibial footprint remaining intact, and anticipated reparability using the BEAR technique. We thus measured interobserver variability in a multiobserver study using intraoperatively obtained arthroscopy videos.

METHODS

Our study was part of an NIH-funded multicenter randomized clinical trial of BEAR (ClinicalTrials.gov identifier: NCT03776162). Institutional review board approval was granted through the Cleveland Clinic. Before patient enrollment, the 11 orthopaedic surgeon investigators—each fellowship trained in sports medicine with extensive experience in ACL reconstruction—watched a short instructional video on how to (1) measure the remaining tibial stump, (2) estimate the proportion of the tibial footprint that remained intact, (3) assess the tibial stump for superficial vascularity, and (4) approximate how many individual tibial stump fiber bundles could be sutured into a unified stump. For purposes of the study, an injured ACL

was defined as potentially repairable if the remaining tibial stump was ≥ 10 mm in length, the remaining tissue could adequately hold a suture, and at least 50% of the tibial footprint remained in continuity with the 10-mm stump.

To assess the variability in ACL rupture pathology in preparation for the trial, these surgeons had identified all ACL reconstructions in their respective practices for at least 6 preceding months, and 5 (K.P.S., L.D.F., P.S., D.F., C.C.K.) had recorded 75 arthroscopic ACL reconstructions. The resulting video clips were trimmed to ≤ 30 seconds and edited to show a concise arthroscopic evaluation of the ACL using a graded arthroscopic probe or an arthroscopic ruler. This 75-patient video library was used for an interrater agreement study by the 11 surgeons. Each video was classified as being of higher or lower technical quality for characterizing the injury based on surgical views shown and imaging resolution. Each surgeon reviewed each video and, based on that review, completed the 6-question ACL Pathology Evaluation Form (Figure 1). These forms were completed between August 20, 2018, and March 30, 2019.

Statistical Analysis

The exact agreement proportion was calculated for all items (ie, the fraction of all paired comparisons of surgeons' ratings of individual patients in which the 2 surgeons chose the same category). In addition, we computed several "chance-corrected" observer agreement measures:

- Fleiss κ and Gwet AC1 (agreement coefficient)^{6,8} statistics for reparability (item 1), gross tibial stump vascularity (item 6), and the dichotomies obtained from splitting the tibial stump length and footprint responses at their respective 10-mm and 50% reparability thresholds
- Respective quadratically weighted ordinal versions of these statistics (κ and AC2) that, for each 4-category ordinal item, penalize adjacent category disagreements only one-ninth and for 2-category

*Address correspondence to Kurt P. Spindler, MD, Department of Orthopaedic Surgery, Orthopaedic and Rheumatologic Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH 44195 (email: SPINDLK@ccf.org; stojsab@ccf.org).

[†]All authors are listed in the Authors section at the end of this article.

The content of this article is solely the responsibility of the authors and does not necessarily represent official views of the National Institutes of Health. Final revision submitted May 8, 2020; accepted May 26, 2020.

One or more of the authors has declared the following potential conflict of interest or source of funding: This project was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (grant R01 1R07131). C.L.C. has a family member who is employed by Smith & Nephew. L.D.F. has received consulting fees from Zimmer Biomet and hospitality payments from the Musculoskeletal Transplant Foundation. P.F. has received education payments from Arthrex. D.F. has received education payments from CDC Medical and Zimmer Biomet; consulting fees from Aastron Biosciences, Ceterix, Conmed Linvatec, DePuy, Hyalex, KCRN, Mitek, MTF, Moximed, Smith & Nephew, Vericel, and Zimmer Biomet; nonconsulting fees from Smith & Nephew and Vericel; faculty speaking fees from Linvatec; and honoraria from Vericel. M.H. has received education payments from Arthrex, nonconsulting fees from Smith & Nephew, and hospitality payments from Kairos Surgical. P.B.I. has received consulting fees from Colgate Palmolive and GE HealthCare. C.C.K. has received grant support from DJO, education payments from CDC Medical, consulting fees from Zimmer Biomet, and nonconsulting fees from Arthrex. B.D.O. has received consulting fees from Conmed/MTF, DePuy/Medical Device Business Services, Linvatec, Musculoskeletal Transplant Foundation, and Rotation Medical; has received royalties from Conmed/MTF; has received honoraria from Vericel; and is a paid associate editor for *The American Journal of Sports Medicine*. P.S. has received education payments from Arthrex and Rock Medical, consulting fees from Arthrex and DJO, and nonconsulting fees from Arthrex. R.W. has stock/stock options in Responsive Arthroscopy. Y.-M.Y. has received education payments from Kairos Surgical and consulting fees from Smith & Nephew. K.P.S. has received research support from Smith & Nephew; has received consulting fees from Flexion Therapeutics, Mitek, the National Football League, Novopeds, and Samumed; has received hospitality payments from Biosense Webster and DePuy; and is on the scientific advisory board for Cytori. DJO provides braces for the ongoing BEAR-MOON study. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval for this study was obtained from Cleveland Clinic (study No. 17-1737).

Case _____
Limit your answer to one response per item:
Item 1: Is the ACL repairable utilizing the BEAR technique?
<input type="checkbox"/> Yes
<input type="checkbox"/> No
Item 2: How long is the tibial stump of the ACL?
<input type="checkbox"/> ≤5 mm
<input type="checkbox"/> 6-9 mm
<input type="checkbox"/> 10-14 mm
<input type="checkbox"/> ≥15 mm
<input type="checkbox"/> Not applicable
Item 3: What percentage of the tibial footprint remains intact?
<input type="checkbox"/> 0-24%
<input type="checkbox"/> 25-49%
<input type="checkbox"/> 50-74%
<input type="checkbox"/> ≥75%
<input type="checkbox"/> Not applicable
Item 4: How long is the femoral stump of the ACL?
<input type="checkbox"/> ≤5 mm
<input type="checkbox"/> 6-9 mm
<input type="checkbox"/> 10-14 mm
<input type="checkbox"/> ≥15 mm
<input type="checkbox"/> Not applicable
Item 5: How many individual fiber bundles could be sutured together and approximated with the BEAR scaffold?
<input type="checkbox"/> 1
<input type="checkbox"/> 2
<input type="checkbox"/> 3
<input type="checkbox"/> ≥4
<input type="checkbox"/> Not applicable
Item 6: Is there gross vascularity present on/in the tibial stump of the ACL?
<input type="checkbox"/> Yes
<input type="checkbox"/> No

Figure 1. The ACL Pathology Evaluation Form. ACL, anterior cruciate ligament; BEAR, bridge-enhanced ACL repair.

disagreements only four-ninths as severely as disagreements between the highest and lowest categories

Kappa statistics and Gwet's more recently proposed AC statistics⁶ correct for "chance agreement" from different perspectives. Kappa statistics define chance agreement based only on each observer's overall classification fractions as what would occur on average if observers randomly reshuffled their assignments of patients to categories. Kappa statistics thus condition upon the marginal distributions of patients to item categories for each observer. In contrast, the Gwet AC statistics are conditioned upon the overall marginal distribution of response categories across the population of observers and patients, discarding a fraction of observed agreement attributed to essentially random guessing with equal category probabilities by at least

1 observer. These statistics give some credit to marginal agreement; that is, similar departures of observers' overall classification proportions from random equiprobable guessing, even without agreement on individual classifications, seem to behave reasonably with highly concentrated or very disparate observer marginal distributions of patients, for which κ statistics can behave paradoxically or unduly pessimistically.

We assessed κ statistics in the context of published interpretive categories that are often applied in the health science literature,⁸ and we assessed κ and Gwet AC statistics using an alternative method proposed by Gwet, adapted to the same interpretive categories.⁶ Calculations were performed using the *ragree* R functions attributed to Gwet in the GitHub repository, *ragree/R/gwet_agree.coeff2.r*. These functions exclude uncategorized ("NA") responses from the calculations.

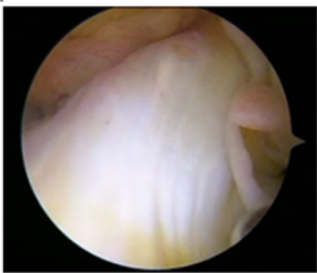
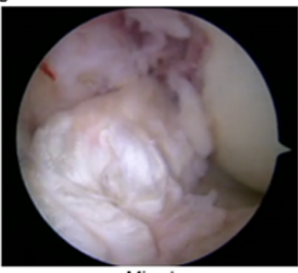
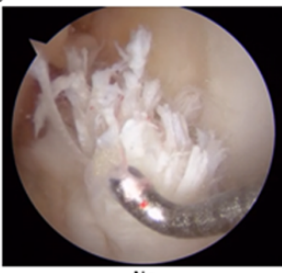



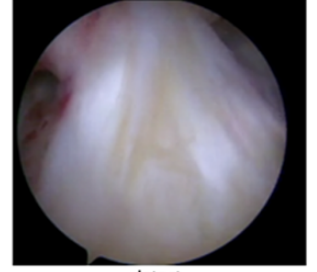
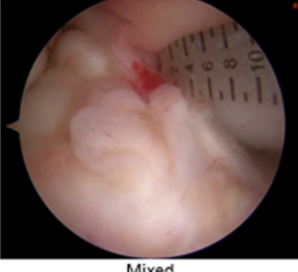
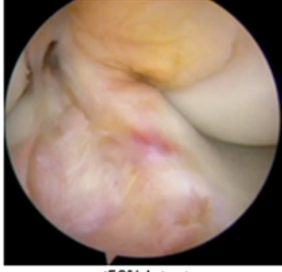
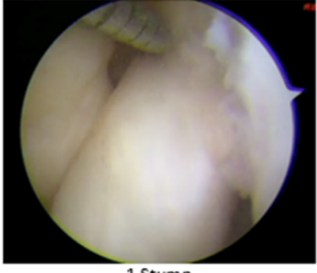
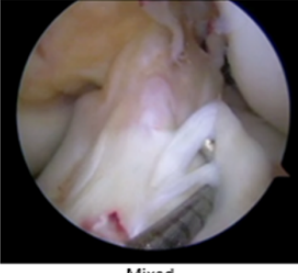

Question	AGREEMENT		
	100%	Mixed	100%
Repair BEAR (Yes or No)	<p>A</p>  <p>Yes</p>	<p>B</p>  <p>Mixed</p>	<p>C</p>  <p>No</p>
Tibial Stump Length	<p>D</p>  <p>>1cm</p>	<p>E</p>  <p>Mixed</p>	<p>F</p>  <p><1 cm</p>
Percent Tibial Footprint Intact	<p>G</p>  <p>Intact</p>	<p>H</p>  <p>Mixed</p>	<p>I</p>  <p><50% Intact</p>
Number Stumps in Repair	<p>J</p>  <p>1 Stump</p>	<p>K</p>  <p>Mixed</p>	<p>L</p>  <p>>1 Stump</p>

Figure 2. Representative arthroscopic images of ACL stumps depicting reparability, tibial stump length, proportion of tibial footprint remaining, and number of stumps capable of being incorporated with a stitch. (A) 100% agreement that ACL can be repaired using BEAR technique; (B) mixed agreement that ACL can be repaired using BEAR technique; (C) 100% agreement that ACL cannot be repaired using BEAR technique; (D) 100% agreement that tibial stump exceeds 1 cm; (E) mixed agreement regarding length of tibial stump; (F) 100% agreement that tibial stump length is <1 cm; (G) 100% agreement that tibial footprint is entirely intact; (H) mixed agreement regarding proportion of tibial footprint remaining intact; (I) 100% agreement that tibial footprint is <50% intact with the stump; (J) 100% agreement that a single stump is present; (K) mixed agreement regarding number of stumps capable of holding a suture; (L) 100% agreement that >1 stump capable of holding a suture is present. ACL, anterior cruciate ligament; BEAR, bridge-enhanced ACL repair.

RESULTS

Of the 75 videos, 50 were judged to be of high quality. All 11 surgeons reviewed the videos on a personal computer and returned the ACL Pathology Evaluation Form. Figure 2

provides representative images taken from the arthroscopy video catalogue used in this study, demonstrating each variable assessed by the raters, with examples of ACL ruptures for which there was high agreement (in a positive or negative sense) or mixed agreement among the observers. Table

TABLE 1
Pooled Responses to ACL Pathology Classification Questions^a

Question	Yes	No			
Reparability	627 (76.0)	198 (24.0)			
Tibial stump vascularity	592 (71.8)	230 (27.9)			
Stump length, mm	≤5	6-9	10-14	≥15	NA
Femoral	539 (65.3)	152 (18.4)	47 (5.7)	33 (4.0)	54 (6.5)
Tibial	46 (5.6)	122 (14.8)	261 (31.6)	387 (46.9)	9 (1.1)
	0%-24%	25%-49%	50%-74%	75%-100%	NA
Tibial footprint intact	30 (3.6)	61 (7.4)	153 (18.5)	546 (66.2)	35 (4.2)
	1	2	3	≥4	NA
No. of suturable fiber bundles	404 (49)	300 (36.4)	60 (7.3)	42 (5.1)	19 (2.3)

^aValues are presented as No. (%). ACL, anterior cruciate ligament; NA, uncategorized responses.

TABLE 2
Observed Exact and Quadratically Weighted Fleiss κ and Gwet AC Statistics for ACL Tear Pathology Classifications^a

Variable	Observed Agreement		Chance-Corrected Agreement (95% CI)	
	Exact	Quadratically Weighted	Fleiss κ (Dichotomy) ^b	Gwet AC1 (Dichotomy) ^c
Reparability	0.825	—	0.522 (0.408-0.635)	0.725 (0.625-0.825)
Dichotomy				
Tibial footprint remaining	0.854	—	0.291 (0.207-0.375)	0.816 (0.739-0.893)
Tibial stump length	0.789	—	0.367 (0.243-0.470)	0.687 (0.602-0.771)
Gross tibial stump vascularity	0.731	—	0.331 (0.233-0.429)	0.540 (0.436-0.662)
No. of fiber bundles	0.604	0.925	0.510 (0.403-0.617)	0.822 (0.767-0.877)
Ordinal				
Tibial footprint remaining	0.637	0.920	0.428 (0.360-0.497)	0.851 (0.791-0.911)
Tibial stump length	0.470	0.876	0.155 (0.080-0.231)	0.672 (0.580-0.763)
Femoral stump length	0.558	0.701	0.090 (0.026-0.155)	0.458 (0.360-0.557)

^aAC, agreement coefficient; ACL, anterior cruciate ligament. Dashes indicate data cannot be calculated.

^bOr quadratically weighted κ (ordinal).

^cOr quadratically weighted AC2 (ordinal).

1 shows frequency distributions of the 11 × 75 = 825 responses to each item, pooled over patients and surgeons. Over three-quarters of responses were above the intended thresholds for tibial stump length and tibial footprint for use of the BEAR repair technique (78.5% and 84.7%, respectively), and 85.3% of responses reported either 1 or 2 suturable bundles. Of the responses, 76% viewed the injury as reparable.

Table 2 lists observed agreement as well as point estimates and 95% CIs for chance-corrected agreement measures for each question. Observed exact agreement ranged from 73.1% to 82.5% for the dichotomies studied and from 47.0% to 63.7% for the ordinally categorized quantitative variables, for which quadratically weighted agreement ranged from 70.1% to 92.5%, reflecting relative infrequency of disagreement beyond adjacent categories. Each chance-corrected measure was significantly above zero, most by wide margins. Observed agreement on reparability was 82.5%, slightly higher (85.4%) for the contributing tibial footprint dichotomy

TABLE 3
Kappa Statistic Interpretive Criteria^a

κ	Agreement Characterization
<0.0	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

^aFrom Landis and Koch.⁹

and slightly lower (78.9%) for the contributing tibial stump dichotomy. Observed agreement was 60.4% on the number of tibial bundles and 92.5% when quadratically weighted, reflecting the rarity of disagreement by >1 bundle.

The conventional Landis and Koch⁹ taxonomy and the Gwet⁶ modified approach to its use, which takes variability

of the estimated coefficients into consideration, agree that the κ statistics showed

- Moderate chance-corrected agreement for reparability ($\kappa = 0.52$), quadratically weighted number of fiber bundles ($\kappa = 0.51$), and remaining tibial footprint ($\kappa = 0.43$)
- Fair agreement for gross tibial stump vascularity ($\kappa = 0.33$) and the dichotomies of tibial footprint ($\kappa = 0.29$) and tibial stump length ($\kappa = 0.37$)
- Slight chance-corrected agreement for quadratically weighted remaining femoral ($\kappa = 0.09$) and tibial ($\kappa = 0.16$) stump lengths (Table 3)

Somewhat in contrast, applying the Landis and Koch taxonomy directly to the more lenient agreement coefficients by Gwet classifies agreement as

- Almost perfect for the 4-level and dichotomous categorizations of the tibial footprint remaining and for the number of suturable fiber bundles
- Substantial for both categorizations of tibial length remaining and for reparability
- Moderate for the femoral length remaining and for gross tibial stump vascularity

The somewhat more conservative interpretive approach by Gwet classifies chance-corrected agreement as fair for femoral stump length, moderate for tibial stump length and vascularity, and substantial for other variables.

As a sensitivity analysis, the various measures of agreement were recalculated in the subsample of 50 high-quality scans. Because few measures and no implications substantially differed in the subsample from the results of the full sample, we report only the latter.

DISCUSSION

In this study, 11 fellowship-trained orthopaedic sports medicine surgeons who have each performed >1000 ACL reconstructions in their respective careers, characterized each of 75 arthroscopy videos through the lens of extensive experience performing knee arthroscopy. This study demonstrates that highly experienced fellowship-trained sports medicine surgeons can characterize ACL tears somewhat reproducibly but far from perfectly across institutions, even when relying on video alone without tactile feedback from probing. The rates of agreement in this study were similar to previous interrater agreement studies focusing on topics such as meniscal pathology and articular cartilage lesions.^{4,12}

A recent study evaluated the inter- and intraobserver reliability of a magnetic resonance imaging (MRI)—based classification system of ACL tear “types,” relying exclusively on where the ligamentous disruption occurred (proximal avulsion, proximal, midsubstance, distal, or distal avulsion).²¹ While the authors concluded that this classification system was reliable and reproducible, the proposed classification system did not assess remnant ACL tissue quality or whether the tibial footprint was partially

disrupted: 2 variable factors that are hypothesized to be needed for successful BEAR. Previous interrater studies have utilized similar techniques to our study, with video recording of arthroscopy for evaluation of meniscal or chondral lesions, but to our knowledge, no previous studies have attempted to develop an arthroscopic ACL tear grading system.^{2,4,5,12}

We used multiple measures to describe the extent of agreement among the 11 participating surgeons. Of these, observed agreement measures are most directly consequential to patients. However, some observed agreement will occur even in the presence of random guessing, exaggerating the extent to which observations are reproducible. Chance-corrected measures attempt to remove this exaggeration, but “random guessing” can have multiple meanings. Kappa statistics presume that random guessing is equivalent to redistributing patients across response categories in the same category proportions that each observer has used but without consideration of any patient characteristics. However, since observer recognition or presumption of dominant categories in a sample can produce quite high levels of such chance agreement, κ statistics can be low with little practical consequence when observed agreement is high. The Gwet AC1 and AC2 statistics presume, in contrast, that random guessing is allocation of patients to categories not only without consideration of their characteristics but with equal category probabilities, thereby giving credit for observers who agree on overall category frequencies although not on individual patients. But the level of reproducibility suggested by these measures may not be exportable to other populations of patients and observers, where different response distributions may apply. For the current data, the latter interpretation of “chance agreement” results in substantially more favorable values than the κ statistics. Interpretation is not straightforward because characterizations designed for interpretation of κ statistics are not necessarily appropriate for interpreting the AC1 and AC2 statistics, and a method proposed by Gwet for incorporating sample variability in the application of interpretive criteria has not been widely adopted.

While not themselves in complete agreement, the current multiple assessments of observer reproducibility do, in aggregate, reveal some clear messages. Pairs of surgeon participants in this study evaluating the same injury agreed roughly 80% of the time on whether (1) at least 50% of the tibial footprint remained, (2) the remaining tibial stump was ≥ 10 mm, and (3) the injury was therefore reparable by the BEAR procedure. In addition, they agreed roughly 60% of the time on exactly how many suturable bundles were available. These characteristics are believed to be most important, among those studied, in determining whether a torn ACL is amenable to repair using the novel BEAR technique. Exact agreement inevitably declines with number of response categories, but disagreements were quite uncommon beyond 1 category in number of fiber bundles, fraction of remaining tibial footprint, or length of remaining tibial stump, as reflected by high observed quadratically weighted agreement statistics, which penalize disagreement proportionally to the square of the number

of intervening categories. Thus, the clinical impacts of differences in surgeon differences in assessments—for enrollment and, conceivably, for covariate adjustment in the BEAR-MOON trial—appear modest and manageable.

Agreement above chance expectation was always statistically significant, whichever measure of this was employed. When chance agreement was based on a surgeon's individual rather than overall uses of the measurement categories (κ statistics), agreement above chance was no more than moderate, a not uncommon finding with κ statistics when observed agreement is high but one that nevertheless reveals clear room for improvement in standardization. The higher values of the Gwet agreement coefficients reflect reductions in expected chance agreement owing to the nonuniformity of overall observed fractions across each item category. But such nonuniformity at least partly stems from using highly trained and experienced surgeons, who will likely have anticipated similar distributions in this broadly inclusive reproducibility study. Thus, the Gwet statistics may somewhat overestimate the reproducibility of these measures in other populations or subpopulations of surgeons and patients.

One major weakness of the current study is the inability to utilize the tactile feedback on which all trained arthroscopic surgeons rely when assessing intra-articular pathology. This study describes agreement for characterizations of patients relying exclusively on visual inspection. Because the participants were given only videos, they were unable to rely on tactile feedback from probing to better assess the injured ligament. This limitation may have affected the study results. In the actual conduct of the BEAR-MOON trial, assessment of suitability for the BEAR approach will be made intraoperatively before randomization, a more information-rich environment allowing custom visualization based on surgeon preference and the benefit of tactile feedback. Although one cannot be sure that surgeons will choose to access and interpret this information similarly, the availability of additional information at surgeon preference could substantially improve reproducibility, particularly if guided by an initial exploratory protocol. It is also worth noting that, although gross tibial stump vascularity is not a prerequisite for the BEAR-MOON procedure, we felt that the presence or absence of stump vascularity may affect biological healing and/or outcome. Thus, the decision was made to include stump vascularity in the ACL Pathology Evaluation Form so that future studies may have the opportunity to assess the effect of vascularity on tissue healing and/or outcomes.

Another limitation of the current study is the inability to correlate the classification system with clinical outcomes. However, the absence of evidence correlating ACL tear “type” with clinical outcomes is at least in part due to the lack of a validated ACL tear classification schema. The current study represents the first step toward achieving this ultimate goal, and correlating ACL tear “type” with clinical outcomes represents one avenue for further study. Likewise, correlating the current ACL classification system with MRI findings, potentially creating an equally reliable

and accurate MRI-based classification system, is another area of future research.

CONCLUSION

This study is the first of its kind to demonstrate the interobserver reliability of arthroscopic classification of ACL tears. We have demonstrated that this classification system, although not ideally reproducible, is reliable enough across surgeons at multiple institutions for use in multicenter studies.

ACKNOWLEDGMENT

The authors thank Brittany Stojavljevic, Cleveland Clinic Foundation, for editorial management.

AUTHORS

BEAR-MOON: José F. Vega, MD (Department of Orthopaedics, Cleveland Clinic Foundation, Cleveland, Ohio, USA); Gregory J. Strnad, MS (Department of Orthopaedics, Cleveland Clinic Foundation, Cleveland, Ohio, USA); Isaac Briskin, MA (Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio, USA); Charles L. Cox, MD, MPH (Vanderbilt University Medical Center, Nashville, Tennessee, USA); Lutul D. Farrow, MD (Department of Orthopaedics, Cleveland Clinic Foundation, Cleveland, Ohio, USA); Paul Fadale, MD (Department of Orthopaedic Surgery, Alpert Medical School of Brown University, Providence, Rhode Island, USA); David Flanigan, MD (Department of Orthopaedics, The Ohio State University, Columbus, Ohio, USA); Michael Hulstyn, MD (Department of Orthopaedic Surgery, Alpert Medical School of Brown University, Providence, Rhode Island, USA); Peter B. Imrey, PhD (Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio, USA); Christopher C. Kaeding, MD (Department of Orthopaedics, The Ohio State University, Columbus, Ohio, USA); Brett D. Owens, MD (Alpert Medical School of Brown University, Providence, Rhode Island, USA); Paul Saluan, MD (Department of Orthopaedics, Cleveland Clinic Foundation, Cleveland, Ohio, USA); Rick Wright, MD (Department of Orthopaedic Surgery, Vanderbilt University Medical Center, Nashville, Tennessee, USA); Yi-Meng Yen, MD, PhD (Division of Sports Medicine, Department of Orthopaedics, Boston Children's Hospital, Boston, Massachusetts, USA); and Kurt P. Spindler, MD (Department of Orthopaedics, Cleveland Clinic Foundation, Cleveland, Ohio, USA).

REFERENCES

1. Ahmad SS, Schürholz K, Liechti EF, Hirschmann MT, Kohl S, Klenke FM. Seventy percent long-term survival of the repaired ACL after dynamic intraligamentary stabilization. *Knee Surg Sports Traumatol Arthrosc.* 2020;28(2):594-598.
2. Anderson AF, Irrgang JJ, Dunn W, et al. Interobserver reliability of the International Society of Arthroscopy, Knee Surgery and Orthopaedic

- Sports Medicine (ISAKOS) classification of meniscal tears. *Am J Sports Med.* 2011;39(5):926-932.
3. Biercevicz AM, Proffen BL, Murray MM, Walsh EG, Fleming BC. T2* relaxometry and volume predict semi-quantitative histological scoring of an ACL bridge-enhanced primary repair in a porcine model. *J Orthop Res.* 2015;33(8):1180-1187.
 4. Dunn WR, Wolf BR, Amendola A, et al. Multirater agreement of arthroscopic meniscal lesions. *Am J Sports Med.* 2004;32(8):1937-1940.
 5. Fresneda MJ, Dere JJ, Yacuzzi CH, Paz MC. ISAKOS classification of meniscal tears: intra and interobserver reliability. *Orthop J Sports Med.* 2015;2(4)(suppl):2325967114S00239.
 6. Gwet KL. Chapter 6: Benchmarking inter-rater reliability coefficients. In: *Handbook of Inter-rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters.* 4th ed. Advanced Analytics, LLC; 2014:163-182.
 7. Herzog MM, Marshall SW, Lund JL, Pate V, Mack CD, Spang JT. Trends in incidence of ACL reconstruction and concomitant procedures among commercially insured individuals in the United States, 2002-2014. *Sports Health.* 2018;10(6):523-531.
 8. Kaeding CC, Pedroza AD, Reinke EK, Huston LJ, Spindler KP. Risk factors and predictors of subsequent ACL injury in either knee after ACL reconstruction: prospective analysis of 2488 primary ACL reconstructions from the MOON cohort. *Am J Sports Med.* 2015;43(7):1583-1590.
 9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
 10. Mahapatra P, Horriat S, Anand BS. Anterior cruciate ligament repair—past, present and future. *J Exp Orthop.* 2018;5(1):20.
 11. MARS Group; Magnussen RA, Borchers JR, et al. Risk factors and predictors of significant chondral surface change from primary to revision anterior cruciate ligament reconstruction: a MOON and MARS cohort study. *Am J Sports Med.* 2018;46(3):557-564.
 12. Marx RG, Connor J, Lyman S, et al. Multirater agreement of arthroscopic grading of knee articular cartilage. *Am J Sports Med.* 2005;33(11):1654-1657.
 13. McIntyre V, Hopper GP, Mackay GM. Anterior cruciate ligament repair in a professional soccer player using internal brace ligament augmentation: a case report focusing on rehabilitation. *Surg Technol Int.* 2019;35:341-348.
 14. Murray MM, Flutie BM, Kalish LA, et al. The bridge-enhanced anterior cruciate ligament repair (BEAR) procedure: an early feasibility cohort study. *Orthop J Sports Med.* 2016;4(11):2325967116672176.
 15. Murray MM, Kalish LA, Fleming BC, et al. Bridge-enhanced anterior cruciate ligament repair: two-year results of a first-in-human study. *Orthop J Sports Med.* 2019;7(3):2325967118824356.
 16. Nwachukwu BU, Patel BH, Lu Y, Allen AA, Williams RJ. Anterior cruciate ligament repair outcomes: an updated systematic review of recent literature. *Arthroscopy.* 2019;35(7):2233-2247.
 17. Perrone GS, Proffen BL, Kiapour AM, Sieker JT, Fleming BC, Murray MM. Bench-to-bedside: bridge-enhanced anterior cruciate ligament repair. *J Orthop Res.* 2017;35(12):2606-2612.
 18. Proffen BL, Perrone GS, Roberts G, Murray MM. Bridge-enhanced ACL repair: a review of the science and the pathway through FDA investigational device approval. *Ann Biomed Eng.* 2015;43(3):805-818.
 19. Ramkumar PN, Hadley MD, Jones MH, Farrow LD. Hamstring autograft in ACL reconstruction: a 13-year predictive analysis of anthropometric factors and surgeon trends relating to graft size. *Orthop J Sports Med.* 2018;6(6):2325967118779788.
 20. Sanders TL, Maradit Kremers H, Bryan AJ, et al. Incidence of anterior cruciate ligament tears and reconstruction: a 21-year population-based study. *Am J Sports Med.* 2016;44(6):1502-1507.
 21. van der List JP, Mintz DN, DiFelice GS. The location of anterior cruciate ligament tears: a prevalence study using magnetic resonance imaging. *Orthop J Sports Med.* 2017;5(6):2325967117709966.
 22. Wright RW, Huston LJ, Haas AK, et al. Effect of graft choice on the outcome of revision anterior cruciate ligament reconstruction in the Multicenter ACL Revision Study (MARS) cohort. *Am J Sports Med.* 2014;42(10):2301-2310.