Multirater Agreement of the Causes of Anterior Cruciate Ligament Reconstruction Failure

A Radiographic and Video Analysis of the MARS Cohort

Matthew J. Matava,*[†] MD, Robert A. Arciero,[‡] MD, Keith M. Baumgarten,[§] MD, James L. Carey,[||] MD, Thomas M. DeBerardino,[‡] MD, Sharon L. Hame,[¶] MD, Jo A. Hannafin,[#] MD, PhD, Bruce S. Miller,** MD, Carl W. Nissen,[††] MD, Timothy N. Taft,[‡‡] MD, Brian R. Wolf,[§§] MD, Rick W. Wright,[†] MD, and The MARS Group[||||]
Investigation performed at Washington University, St Louis, Missouri, USA

Background: Anterior cruciate ligament (ACL) reconstruction failure occurs in up to 10% of cases. Technical errors are considered the most common cause of graft failure despite the absence of validated studies. Limited data are available regarding the agreement among orthopaedic surgeons regarding the causes of primary ACL reconstruction failure and accuracy of graft tunnel placement.

Hypothesis: Experienced knee surgeons have a high level of interobserver reliability in the agreement about the causes of primary ACL reconstruction failure, anatomic graft characteristics, and tunnel placement.

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

Methods: Twenty cases of revision ACL reconstruction were randomly selected from the Multicenter ACL Revision Study (MARS) database. Each case included the patient's history, standardized radiographs, and a concise 30-second arthroscopic video taken at the time of revision demonstrating the graft remnant and location of the tunnel apertures. All 20 cases were reviewed by 10 MARS surgeons not involved with the primary surgery. Each surgeon completed a 2-part questionnaire dealing with each surgeon's training and practice, as well as the placement of the femoral and tibial tunnels, condition of the primary graft, and the surgeon's opinion as to the causes of graft failure. Interrater agreement was determined for each question with the kappa coefficient and the prevalence-adjusted, bias-adjusted kappa (PABAK).

Results: The 10 reviewers have been in practice an average of 14 years and have performed at least 25 ACL reconstructions per year, and 9 were fellowship trained in sports medicine. There was wide variability in agreement among knee experts as to the specific causes of ACL graft failure. When participants were specifically asked about technical error as the cause for failure, interobserver agreement was only slight (PABAK = 0.26). There was fair overall agreement on ideal femoral tunnel placement (PABAK = 0.55) but only slight agreement on whether a femoral tunnel was too anterior (PABAK = 0.24) and fair agreement on whether it was too vertical (PABAK = 0.46). There was poor overall agreement for ideal tibial tunnel placement (PABAK = 0.17).

Conclusion: This study suggests that more objective criteria are needed to accurately determine the causes of primary ACL graft failure as well as the ideal femoral and tibial tunnel placement in patients undergoing revision ACL reconstruction.

Keywords: revision; anterior cruciate ligament; tunnel placement; interobserver reliability

Anterior cruciate ligament (ACL) tears are a common cause of disability in patients involved in cutting, pivoting, and jumping activities. It is estimated that approximately 200,000 ACL reconstructions are performed in the United States each year in an attempt to restore knee stability and return patients to an active lifestyle for both work and recreational activities.[3,5,8,35] Unfortunately, failure of the primary reconstruction has been noted for between 0.7% and 10% of cases,[1,11,16,18,19,24,25,28,33,34] resulting in an estimated 10,000 to 20,000 revision reconstructions performed annually.[15,20,31,33,35]

Understanding the specific cause of primary reconstruction failure is paramount to improving revision ACL reconstruction. It has been widely assumed, based on case series with evidence levels 4 and 5[39] as well as expert opinion,[9,17,18,22] that technical errors[1,12,16,18,23,24,32,34]—in particular femoral tunnel malposition[32,34,36]—are the most

common causes of graft failure. Yet there is a paucity of high-quality studies validating this assertion. A significant challenge to such studies is the limited number of revision ACL reconstructions performed by an individual surgeon or institution. The Multicenter ACL Revision Study (MARS) was conceived as a prospective longitudinal cohort to address predictors and prognosis of revision ACL surgery specifically in regard to activity level, health-related and knee-related quality of life, and future risk of osteoarthritis. This multicenter format has the benefit of significantly increasing the number of patients available so that factors potentially influencing patient outcome can be better evaluated. However, for this multicenter format to be effective, it is imperative that participating surgeons provide reliable and reproducible clinical data, agreeing on the anatomic and technical factors associated with graft placement as well as graft failure. If agreement among surgeons as to the cause of failure is poor, then a more reliable means of assessing the clinical factors responsible for ACL failure is necessary.

It was our hypothesis that experienced knee surgeons who perform revision ACL surgery would have a reasonably high level of agreement as to the cause of the primary reconstruction failure. Consistency in evaluation and documentation of the cause of failure is vital to meaningful analysis of long-term outcomes of treatment. Therefore, the purpose of this study was to determine the interobserver agreement among experienced sports medicine specialists in the MARS group regarding specific technical and anatomic factors of the primary graft as well as the cause of the primary reconstruction failure. Assessments were made based on the patient's history, standardized radiographs, and arthroscopic videos of the failed index surgery. The results of this multicenter trial could be instrumental in determining the factors responsible for primary graft failure as well as the optimal methods of performing revision surgery.

MATERIALS AND METHODS

Demographics of the MARS Group

The MARS surgeon group is composed of voluntary members of the American Orthopaedic Society for Sports Medicine (AOSSM). The majority of surgeons were fellowship trained in sports medicine, and all completed a 6-hour training course outlining the goals of MARS as well as the method by which to complete the MARS Surgeon Form. A letter was sent to all 89 participating MARS surgeons with at least 5 years of clinical experience that described the purposes of the study and asked whether they would be interested in participating as a case reviewer. Of those surgeons who agreed to participate, 10 were randomly selected to act as reviewers.

Data Collection

For the current study, all participating surgeons were queried as to their interest in participating. Those who volunteered were then asked to submit a random revision ACL reconstruction case for analysis. There were no stipulations as to the patient demographics or number of comorbidities. However, a case was not accepted if the primary reconstruction was a double-bundle construct, if the known cause of failure was due to infection, or if the patient had already undergone at least one revision reconstruction.

A total of 20 cases were selected for review representing a variety of failed reconstructions. Each case included 3 basic sets of data: case history, standardized radiographs, and a concise 30-second arthroscopic video demonstrating the primary reconstructive graft (or its remnant) and location of the tibial and femoral tunnel apertures and hardware (if any). Each case history contained the following data: patient age, gender, mechanism of primary ACL tear, date of primary ACL reconstruction, primary graft source, method of fixation, prior surgical technique if known (eg, 1-incision transtibial, 1-incision anteromedial portal, 2-incision, or open reconstruction), and date of revision reconstruction. Standardized MARS radiographic images for each case history included a standing bilateral anteroposterior view in full extension, a lateral view in maximum extension, bilateral flexion weightbearing view at 45° (Rosenberg view), bilateral 45° patellar (Merchant) views, and bilateral standing alignment (hip-knee-ankle) views. The video accompanying each case consisted of a concise, approximately 30-second segment of the arthroscopic video taken during the revision surgery by use of

*Address correspondence to Matthew J. Matava, MD, Department of Orthopedic Surgery, Washington University School of Medicine, 14532 South Outer Forty Drive, Chesterfield, MO 63141, USA (e-mail: matavam@wudosis.wustl.edu).

†Department of Orthopedic Surgery, Washington University School of Medicine, St Louis, Missouri, USA.
‡University of Connecticut Health Center, Farmington, Connecticut, USA.
§Orthopaedic Institute, Sioux Falls, South Dakota, USA.
‖Vanderbilt University, Nashville, Tennessee, USA.
¶University of California–Los Angeles, Los Angeles, California, USA.
#Hospital for Special Surgery, New York, New York, USA.
**University of Michigan Medical School, Ann Arbor, Michigan, USA.
††Connecticut Children's Medical Center, Farmington, Connecticut, USA.
‡‡University of North Carolina, Chapel Hill, North Carolina, USA.
§§University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA.
‖‖All members are listed in the Contributing Authors section at the end of this article.

a 30° arthroscope placed in the anterolateral and antero-medial portals. Each video demonstrated the surgeon prob-ing the failed graft (if present) to assess graft attenuation or absence. The videos also showed the femoral and tibial tunnel apertures and their size once the failed graft had been debrided and hardware removed by the operating surgeon.

A compilation of all 20 case histories, corresponding radiographs, and arthroscopic videos was saved on a digital video disc (DVD) and sent to a random selection of 10 par-ticipating MARS surgeons possessing at least 5 years of clinical experience. Each surgeon was asked to complete a 2-part questionnaire regarding each case (see the Appen-dix, available in the online version of this article at http://ajsm.sagepub.com/supplemental). Part 1 was com-posed of 6 questions that dealt solely with each surgeon's practice type and experience in performing primary and revision ACL surgery. Part 2, composed of 21 questions that required a simple "yes" or "no" answer, was adapted from the 48-page MARS Surgical Form completed for each patient enrolled in the MARS study. These questions were concerned specifically with the nature of the primary graft (ie, absent, present but elongated, torn), placement of the femoral and tibial tunnels (eg, too anterior, too poste-rior, too vertical, etc), and the surgeon's opinion as to the cause of failure (ie, traumatic, biologic, technical, combina-tion). The reviewers were not given any sort of primer or instruction on the objective gold standard or predeter-mined correct response regarding the accuracy of tunnel placement at the inception of the study. None of the 20 cases was submitted by one of the reviewers.

Statistical Analysis

Interrater agreement regarding the various responses on part 2 of the questionnaire was analyzed by use of several different measures. Interrater agreement (with CI) was determined for each question by calculating the percentage of perfect agreement among all pairwise, between-rater comparisons. There were 56 pairwise comparisons. Those comparisons for which 8 of 10 surgeons agreed on the ques-tion of interest were also determined since this number (80%) was felt to represent a reasonably high degree of agreement. Statistical analyses were performed with SAS software, version 9.2 (SAS Institute Inc). This provided a percentage of agreement between raters. In addition, the number of cases in which 8 of 10 reviewers agreed was determined for each question.

The Cohen kappa (κ) coefficient was also calculated to assess interrater agreement. Kappa seeks to express inter-rater agreement beyond that expected by chance alone through the following equation:

$$\kappa = \frac{observed\ agreement - chance\ agreement}{1 - chance\ agreement}.$$

Interpreting κ using this model, however, assumes a gener-ally equal distribution of prevalence of the studied attri-bute. If prevalence is not equally distributed, then the κ

value is distorted and becomes less meaningful. The prev-alence index (PI) was used to determine the appropriate-ness of the κ value and was calculated with the following equation:

$$PI = \frac{No.\ of\ yes - yes\ responses - No.\ of\ no - no\ responses}{Total\ No.\ of\ responses}.$$

The PI ranges from 0 to 1, with a higher PI indicating that κ is less likely to accurately evaluate agreement due to the problems of an uneven prevalence distribution. A prevalence-adjusted, bias-adjusted κ (PABAK) is one method for adjusting κ for the paradoxes caused by large differences between the 2 types of agreement (prevalence) or the 2 types of disagreement (bias).[6] A PABAK is particu-larly useful in cases with high percentage agreement but a low κ coefficient.[10,14,21] A PABAK was calculated using the mean of the observed agreement and disagreement to determine the chance agreement factor. Landis and Koch[27] have provided a commonly used interpretation of κ, where a κ value below 0.0 suggests poor agreement; 0.00 to 0.20, slight agreement, 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agree-ment; and 0.81 to 1.00, almost perfect agreement. This clas-sification was originally developed in the study of agreement between 2 raters, where the κ coefficient reflects error, not low prevalence. This ordinal interpretation scheme was used for the PABAK scores as a descriptor of interobserver agreement. Landis and Koch's categorical nomenclature ("slight agreement," "moderate agreement," etc) was used to provide the reader with easily understood descriptors to aid in interpretation of the numerical values. We chose to apply Landis and Koch's classification to the PABAK to avoid misleading interpretations that reflect the nature of the pop-ulation rather than the observation procedure itself.

RESULTS

The 10 reviewers were in practice an average of 14 years (range, 5-35 years). All performed at least 25 ACL recon-structions per year, and 9 reviewers were fellowship trained in sports medicine. Nine (90%) of the 10 reviewers were in private practice with an academic affiliation. The estimated average number of revision reconstructions per-formed annually by the reviewers was as follows: 4 reviewers with 1-5 revisions, 3 with 6-10 revisions, 1 with 11-15 revisions, 0 with 16-20 revisions, and 2 with >20 revisions.

Overall interobserver agreement (with agreement >80%) is shown in Table 1. The κ, PI, and PABAK values were calculated for each question and are shown in Table 2. Topics with PI values closer to 1.0 indicate a decreased relevance of the κ values and reflect the increased influ-ence of prevalence resulting in larger discrepancies between their κ and PABAK values.

Interobserver agreement for questions regarding the failed graft's presence and condition averaged 87% (range, 83%-90%). At least 80% of the reviewers agreed in 90% (18/

TABLE 1
Interobserver Agreement for Each Radiographic
and Graft-Related Topic Assessed

	Overall Agreement, % (95% CI)	Cases With >80% Agreement, n
Ideal placement		
Femur	77 (75-80)	14
Tibia	58 (55-62)	9
Mean	68	11.5
Femoral position		
Vertical	73 (70-76)	12
Anterior	62 (59-65)	9
Posterior	92 (90-94)	18
Mean	76	15
Tibial position		
Medial	89 (87-91)	16
Lateral	90 (89-92)	19
Anterior	85 (82-87)	18
Posterior	72 (69-75)	13
Mean	84	16.5
Tunnel size		
Femur	89 (87-91)	18
Tibia	74 (71-77)	14
Mean	81	16
Graft condition		
Absent	90 (89-93)	19
Elongated	86 (84-88)	17
Torn	83 (81-86)	18
Mean	87	18.5
Graft fixation		
Femur	92 (90-94)	20
Tibia	96 (95-97)	20
Mean	94	10
Cause of failure		
Trauma	70 (67-73)	11
Biologic	76 (74-79)	15
Technical	63 (60-66)	10
Combined	52 (49-55)	3
Other ligament injury	98 (97-99)	20
Mean	72	11.8

20) of the cases. The percentage agreement among the reviewers regarding the specific cause of graft failure averaged 72% (range, 52%-98%). At least 80% of the reviewers agreed on the cause of graft failure in only 55% of the cases. Interobserver agreement was almost perfect and was the highest regarding whether other ligamentous insufficiency was the primary cause (98%) (PABAK = 0.96) and lowest (slight agreement) if a combination of factors was the likely cause of failure (52%) (PABAK = 0.04). At least 80% of the reviewers agreed in 55% (11/20) of cases when estimating the cause of failure.

The highest rate of agreement pertained to ligamentous insufficiency, with at least 80% of the reviewers agreeing in all 20 cases (100%); the lowest rate of agreement pertained to a combination of factors, with only 15% (3/20) of cases having at least 80% reviewer agreement. When reviewers were specifically asked about technical error as the cause for failure, interobserver agreement was only 63%, with 50% (10/20) of cases having at least 80% reviewer agreement.

There was moderate interobserver agreement of 77% (95% CI, 75%-80%; PABAK = 0.55) regarding whether the femoral tunnel was ideal in placement compared with slight agreement of 58% (95% CI, 55%-62%; PABAK = 0.17) for ideal tibial tunnel placement (Figure 1). At least 80% of the reviewers agreed in 70% (14/20) of the cases for the femoral tunnel placement and size and in 45% (9/20) of the cases for the tibial tunnel placement and size. Further analysis of tunnel placement demonstrated that the percentages of agreement for questions regarding specific femoral tunnel placement (ie, too anterior, too posterior, too vertical) averaged 76% (range, 62%-92%). At least 80% of the reviewers agreed in 65% (13/20). There was almost perfect interobserver agreement when posterior femoral tunnel placement was evaluated (92%) (PABAK = 0.84) but only fair agreement when anterior placement (62%) was assessed (PABAK = 0.24). There was moderate agreement regarding femoral tunnel verticality (mean, 73%; range, 70%-76%; PABAK = 0.46).

The percentage agreement for questions about specific tibial tunnel placement averaged 84% (range, 72%-90%), with at least 80% of the reviewers agreeing on specific tibial tunnel position in 83% (17/20) of the cases. Interobserver agreement was almost perfect and was highest when lateral tibial tunnel placement was evaluated (PABAK = 0.81) and lowest (moderate agreement) when posterior placement was assessed (PABAK = 0.43) (Figure 2).

DISCUSSION

This study demonstrated wide variability in the agreement among experienced knee surgeons when assessing certain key aspects of failed ACL reconstructions. Specifically, significant variability in agreement was found regarding the surmised cause of graft failure and tunnel placement. Overall interobserver agreement was 77% (PABAK = 0.55) when reviewers were determining whether the femoral tunnel was ideal in placement and size, compared with 58% (PABAK = 0.17) agreement for the tibial tunnel. In addition, the most commonly espoused technical cause of primary graft failure—anterior femoral tunnel placement—was agreed upon in only 62% of cases (PABAK = 0.24). Agreement regarding femoral tunnel verticality (another recently recognized cause of technical error) was somewhat better (mean agreement: 73% [PABAK = 0.46]). These results are in agreement with the work of Morgan et al,[32] who analyzed 460 revision ACL reconstructions and cited "technical cause of failure" in 60% of the cases, with femoral tunnel malposition found to be the most commonly cited technical reason for graft failure (48% of the 460 cases). We can only hypothesize that tibial tunnel agreement was worse than femoral tunnel agreement due to the relative importance given to the femoral tunnel in terms of prior basic and clinical research. In other words, more attention has been directed at the femoral tunnel because of its presumed greater perceived importance in graft function compared with the tibial tunnel.

Prior research has attempted to determine interobserver reliability in the assessment of femoral tunnel

TABLE 2
Kappa Coefficient, Prevalence Index, and PABAK Values for Each Radiographic and Graft-Related Topic Assessed[a]

	κ Coefficient	Prevalence Index	PABAK	Landis and Koch[27] Agreement
Ideal placement				
Femur	0.16	0.68	0.55	Moderate
Tibia	0.15	0.13	0.17	Slight
Femoral positioning				
Vertical	0.37	0.40	0.46	Moderate
Anterior	0.24	0.04	0.24	Fair
Posterior	0.18	0.9	0.84	Almost perfect
Tibial positioning				
Medial	0.37	0.81	0.78	Substantial
Lateral	0.16	0.88	0.81	Almost perfect
Anterior	0.32	0.74	0.69	Substantial
Posterior	0.19	0.55	0.43	Moderate
Tunnel size				
Femur	0.20	0.85	0.78	Substantial
Tibia	0.14	0.63	0.48	Moderate
Graft condition				
Absent	0.81	0.066	0.82	Almost perfect
Elongated	0.70	0.28	0.72	Substantial
Torn	0.45	0.63	0.67	Substantial
Graft fixation				
Femur	0.40	0.86	0.84	Almost perfect
Tibia	−0.015	0.96	0.92	Almost perfect
Cause of failure				
Trauma	0.43	0.058	0.40	Fair
Biologic	0.095	0.70	0.53	Moderate
Technical	0.056	0.47	0.26	Fair
Combination	0.056	0.015	0.04	Slight
Other ligament injury	0.10	0.98	0.96	Almost perfect

[a]PABAK, prevalence-adjusted, bias-adjusted kappa.

placement in primary ACL reconstruction. Warme et al[36] prospectively evaluated the postoperative plain radiographs of 54 patients after primary ACL reconstruction. Three blinded reviewers performed 8 different radiographic measurements to assess tunnel location. Intraobserver reliability for femoral measurements ranged from none to substantial but for tibial tunnel measurements was moderate to almost perfect. Interobserver reliability ranged from slight to moderate for femoral measures and from fair to substantial for tibial tunnel measures. In this series, the presence of metal interference screws did not improve the reliability of measurements. These authors concluded that radiographic tunnel measurements after ACL reconstruction are quite variable, with reliability falling only into the fair to moderate categories.
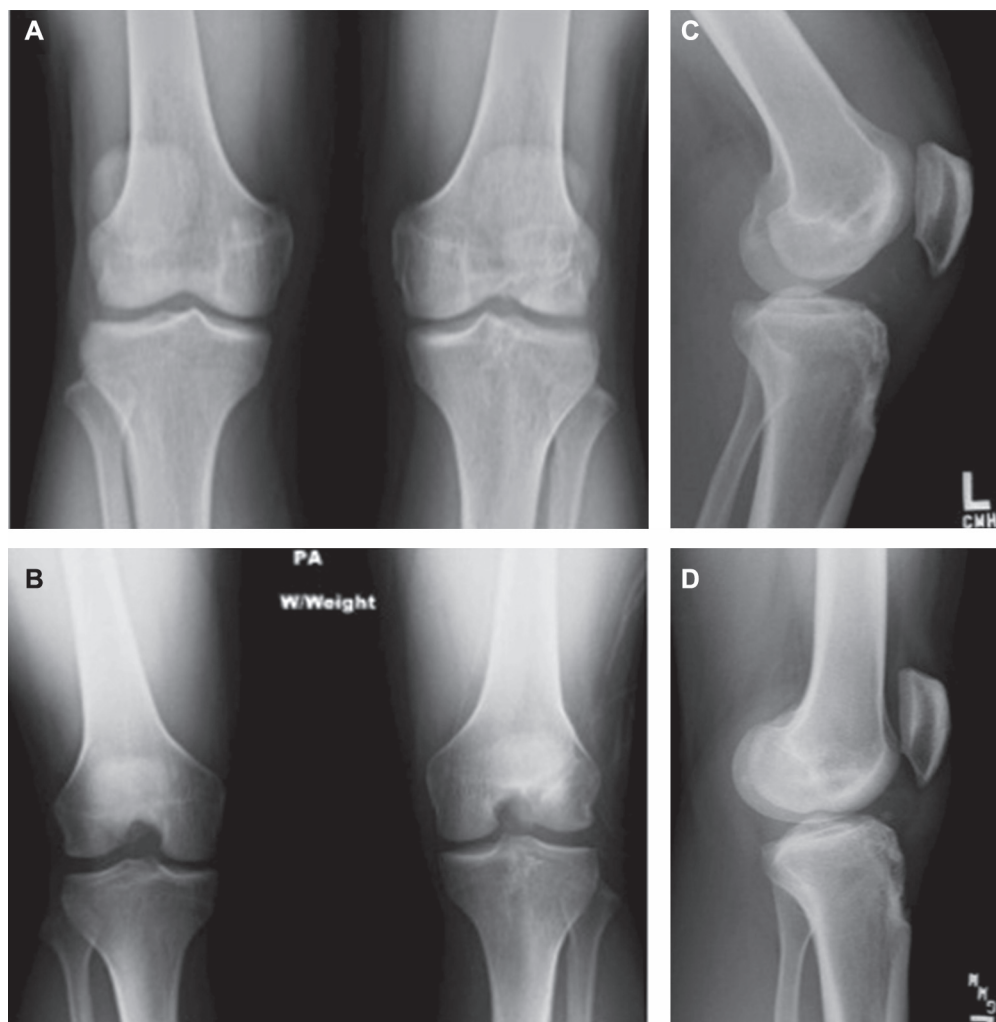
Wolf et al[37] analyzed variation in ACL tunnel placement between surgeons and the influence of preferred surgical technique and surgeon experience using 3-dimensional computed tomography. There was a relatively high degree of intrasurgeon reliability in the placement of ACL graft tunnels. The location of the femoral tunnel aperture in the sagittal plane relative to the notch roof was the most variable measurement, with a range of means of 16%. There was, however, variability of average tunnel placement of up to 22% of the mean condylar depth, likely reflecting the difference in individual surgeons' preferred tunnel locations.

Interestingly, surgeon experience level did not appear to significantly affect tunnel location.

McConkey et al[30] evaluated arthroscopic agreement among surgeons on primary ACL tunnel placement. The investigators found that operating surgeons were more likely to judge their own tunnels more favorably than were other observers. However, independent surgeon reviewers appeared to be more critical of other surgeons' tunnels. The investigators concluded that overall, surgeons do not agree on the ideal placement for single-bundle ACL tunnels.

Multiple studies have assessed the interobserver and intraobserver reliability in the arthroscopic evaluation and classification of other knee pathology.[2,4,7,13,29] Marx et al[29] demonstrated good interobserver reliability with arthroscopic classification of articular cartilage lesions. Six experienced surgeons classified 31 different lesions based on video analysis. The authors reported 81% to 94% agreement depending on lesion location; however, κ values varied between fair to near perfect agreement (range, 0.34-0.87).

Brismar et al[4] studied both intra- and interobserver reliability of arthroscopic classification of mild to moderate osteoarthritis using video assessment. Four different surgeons reviewed 19 different videotaped knee arthroscopies twice, classifying the observed arthritis using the Outerbridge, Collins, and French Society of Arthroscopy measures. The
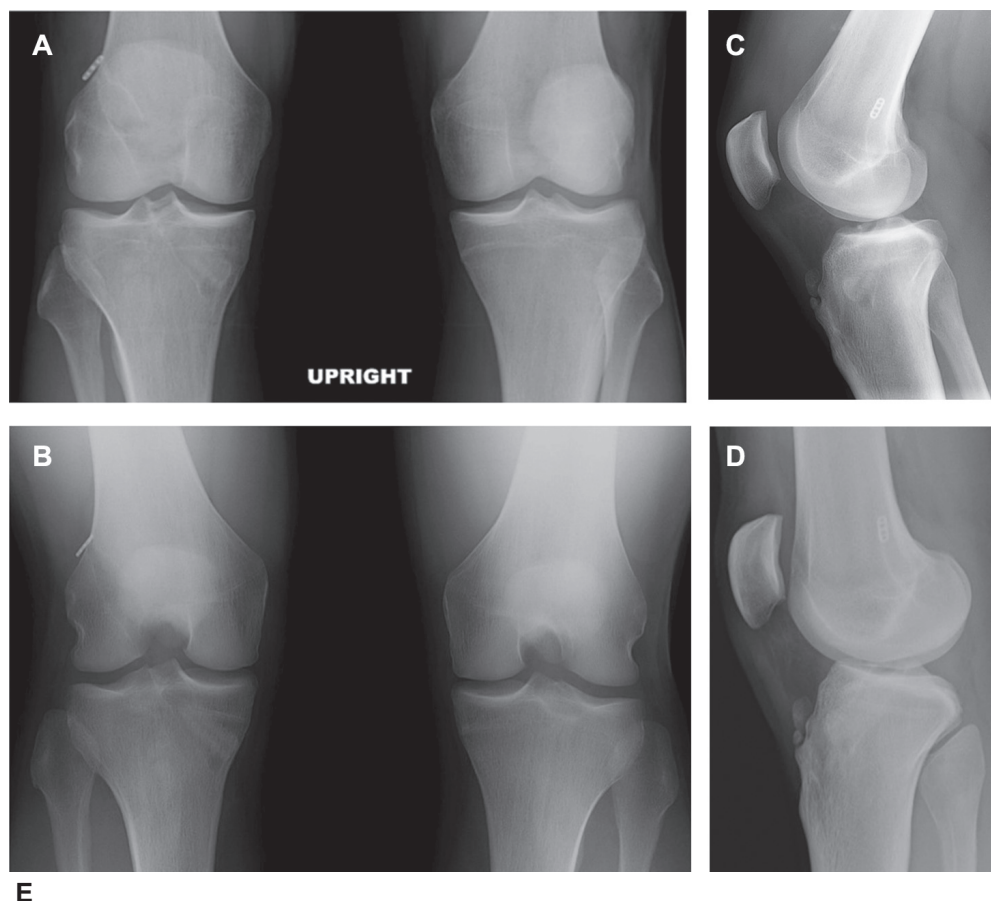
Figure 1. Case 1. Selected radiographic views: (A) weightbearing anteroposterior; (B) 45° flexion–weightbearing (Rosenberg); (C) 30° lateral; and (D) full-extension lateral. (E) Selected questions pertaining to tunnel location and number of corresponding ''yes'' or ''no'' responses.

investigators found 59% to 62% overall interobserver agreement and 55% to 77% intraobserver agreement with κ values indicating moderate agreement. Studies by both Anderson et al[2] and Dunn et al[13] used intraoperative video analysis

to determine the reliability of different surgeons in assessing meniscal tears by location, depth, type of tear, and treatment. Both studies found that grading of meniscal tears was reliable and reproducible. Of interest, Dunn et al[13] also noted

Figure 2. Case 2. Selected radiographic views: (A) weightbearing anteroposterior; (B) 45° flexion–weightbearing (Rosenberg); (C) 30° lateral; and (D) full-extension lateral. (E) Selected questions pertaining to tunnel location and number of corresponding ''yes'' or ''no'' responses.

Question	Yes	No
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel ideal in terms of both position AND size?	3	7
Do you feel that PRIMARY GRAFT FAILURE was due to insufficient FEMORAL FIXATION?	0	10
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel ideal in terms of position, but ENLARGED?	4	6
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO VERTICAL?	4	6
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO ANTERIOR?	1	9
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO POSTERIOR?	0	10
Do you feel that PRIMARY GRAFT FAILURE was due to insufficient TIBIAL FIXATION?	0	10
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of both position AND size?	4	6
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of position, but ENLARGED?	4	6
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO MEDIAL?	0	10
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO LATERAL?	1	9
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO ANTERIOR?	0	10
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO POSTERIOR?	3	7

the effect of the prevalence paradox in their study, with certain categories having low κ values despite high observed agreement. The investigators specifically noted that their conclusions were based on their percentage of agreement rather than κ coefficients due to the problems related to κ.

While the Cohen κ coefficient has been used in multiple reliability studies to assess agreement, κ had limited usefulness in this study due to several well-documented problems. These problems, referred to as "paradoxes," limit the application and interpretation of κ, leading many

statisticians to warn against using κ alone to evaluate agreement. The paradox of prevalence was particularly relevant to this study and can be responsible for significantly depressed κ values despite high observer agreement, such as in the study by Marx et al.[29] This paradox results from a high prevalence of one type of agreement compared with the converse (ie, "yes-yes" vs "no-no" agreement between observers), which causes a significant increase in the chance agreement correction. For example, in this study when reviewers were asked about tibial fixation as the

cause of graft failure, there was 96% overall interobserver agreement. Yet, because the distribution of agreement was substantially uneven (96% of reviewers agreed that tibial fixation was *not* the cause of failure in the 20 cases), κ was significantly decreased and actually resulted in a negative value (–0.0152). The high PI value (0.96), however, forewarns of significant κ distortion. Similar problems due to the prevalence paradox were seen throughout this study. One method of resolving this dilemma is adjusting the κ coefficient by using the mean of the observed agreement and disagreement when calculating the chance agreement factor. This adjustment, referred to as PABAK, eliminates the problem (the κ paradox) caused by uneven distribution of prevalence and bias.[6,10,14] The PABAK values, shown in Figure 2, reflect agreement without the influence of prevalence or bias. Statisticians have warned that similar to κ, PABAK should not be used alone to interpret agreement, as prevalence and bias have informative value in assessing agreement.[10,26] We used the Landis and Koch[27] classification to provide useful benchmarks to interpret agreement. This classification was developed in the study of agreement between 2 raters, where the κ coefficient reflects error, not low prevalence. Due to the low prevalence found in our study and the associated effect on κ, we have reported multiple statistics (ie, κ, prevalence index, and PABAK) so that readers can interpret the adequacy of agreement within their specific context.

Several limitations to this study should be addressed. Although video has been consistently used for reliability studies,[2,4,13,29] it has some limitations. The degree of visual assessment is limited by the quality of what is shown on the video, which is dependent upon the arthroscopic skills of the surgeon as well as the quality of the arthroscopic camera and video software. Additionally, there is no tactile feedback, which is typically achieved with probing and the use of other instrumentation as would be possible had the reviewing surgeons actually performed the surgery themselves. While it is possible that resolution of these factors would further improve reliability in real operative settings, it is conceivable that it could further confound agreement. Additionally, the 30-second video and radiographs may not allow for a detailed preoperative assessment as would be possible in the clinical setting, where physical examination, adjunctive MRI, and/or other imaging studies would be available. Some technical causes of ACL graft failure (ie, tibial tunnel too lateral) are too rare to ensure adequate representation among the cases randomly submitted for review. Having more than 20 cases may have solved this issue, but the number of cases of each potential cause of failure would still likely not be completely representative of all causes of ACL reconstruction failure. Another weakness relates to lack of an objective gold standard or predetermined correct response regarding the accuracy of tunnel placement and the specific cause of graft failure for each case. This may have been addressed by providing the reviewing surgeons a classification system before their analysis of the cases. However, our purpose with this study was not to determine how often these experienced knee surgeons chose the correct answer to the various topics of

interest but rather to discern how often they agreed on various factors associated with the causes of graft failure and the accuracy of graft placement. Finally, the results of this study are potentially limited by the fact that experienced orthopaedic knee surgeons (as shown in this as well as other studies[36-38] discussed above) do not uniformly agree on what constitutes ideal tunnel placement after ACL reconstruction. This objective can only be accomplished if (1) a simple, uniform definition of ideal tunnel location can be agreed upon based on validated anatomic and radiographic landmarks and reference points and (2) this information is widely disseminated and used by surgeons who perform this procedure. This is, perhaps, even more important for inexperienced knee surgeons who infrequently perform ACL reconstruction.

There are several strengths of this study that make it unique compared with prior literature dealing with ACL revision. This is the first study, to our knowledge, to assess the interobserver reliability when evaluating primary ACL reconstruction failure, particularly focusing on graft location, anatomic graft characteristics, and cause of failure. In all cases, the reviewing surgeons used the same clinical histories, radiographs, and videos and made their assessments independently without collaboration. Each case had uniform clinical, radiographic, and arthroscopic video data available for review, which is necessary to discern the myriad factors associated with a failed primary ACL reconstruction. All cases were reviewed by a number of knee surgeons who all had significant clinical experience performing revision ACL surgery. In addition, our statistical analysis was able to compensate for the absence of some known, albeit rare, causes of primary ACL reconstruction failure to more accurately determine true agreement among reviewers despite the bias potentially associated with agreement by chance alone. Finally, the reviewers chosen for this study had significant interest and experience performing ACL revision surgery. While the results we obtained with this group of reviewers may not be representative of the results we might have obtained with less experienced surgeons, we feel justified in using a more experienced group since it is likely that most ACL revisions are performed by more experienced knee surgeons.

In conclusion, there was wide variability in agreement among knee experts as to the specific cause of primary ACL reconstruction failure and the appropriateness of tunnel placement in patients undergoing ACL revision. Interobserver agreement was only slight when the cause of primary graft failure was attributed to technical error despite this being the most commonly theorized cause of failure. There was fair overall agreement on ideal femoral tunnel placement but only slight agreement on whether a femoral tunnel was too anterior and fair agreement on whether it was too vertical. There was poor overall agreement for ideal tibial tunnel placement. This study suggests that more objective measures are needed to accurately determine the cause of primary ACL graft failure as well as ideal tunnel location to improve the outcome of ACL reconstruction and facilitate future research in revision ACL surgery.

CONTRIBUTING AUTHORS

John P. Albright, MD (University of Iowa Hospitals and Clinics, Iowa City, Iowa); Christina R. Allen, MD (University of California–San Francisco, San Francisco, California); Annunziato Amendola, MD (University of Iowa Hospitals and Clinics, Iowa City, Iowa); Allen F. Anderson, MD (Tennessee Orthopaedic Alliance, Nashville, Tennessee); Jack T. Andrish, MD (Cleveland Clinic, Cleveland, Ohio); Christopher C. Annunziata, MD (Commonwealth Orthopaedics & Rehabilitation, Arlington, Virginia); Robert A. Arciero, MD (University of Connecticut Health Center, Farmington, Connecticut); Bernard R. Bach Jr, MD (Rush University Medical Center, Chicago, Illinois); Champ L. Baker III, MD (The Hughston Clinic, Columbus, Georgia); Arthur R. Bartolozzi, MD (3B Orthopaedics, Philadelphia, Pennsylvania; University of Pennsylvania Health System, Philadelphia, Pennsylvania); Keith M. Baumgarten, MD (Orthopedic Institute, Sioux Falls, South Dakota); Jeffery R. Bechler, MD (University Orthopedic Associates LLC, New Brunswick, New Jersey); Jeffrey H. Berg, MD (Town Center Orthopaedic Associates, Reston, Virginia); Geoffrey A. Bernas, MD (State University of New York at Buffalo, Buffalo, New York); Stephen F. Brockmeier, MD (University of Virginia, Charlottesville, Virginia); Robert H. Brophy, MD (Washington University, St Louis, Missouri); Charles A. Bush-Joseph, MD (Rush University Medical Center, Chicago, Illinois); J. Brad Butler V, MD (Orthopedic and Fracture Clinic, Portland, Oregon); John D. Campbell, MD (Bridger Orthopaedic and Sports Medicine, Bozeman, Montana); James L. Carey, MD, MPH (University of Pennsylvania, Philadelphia, Pennsylvania); James E. Carpenter, MD (University of Michigan, Ann Arbor, Michigan); Brian J. Cole, MD (Rush University Medical Center, Chicago, Illinois); Daniel E. Cooper, MD (W.B. Carrell Memorial Clinic, Dallas, Texas); Jonathan M. Cooper, DO (HealthPartners Specialty Clinic, St Paul, Minnesota); Charles L. Cox, MD, MPH (Vanderbilt University, Nashville, Tennessee); R. Alexander Creighton, MD (University of North Carolina Medical Center, Chapel Hill, North Carolina); Diane L. Dahm, MD (Mayo Clinic Rochester, Rochester, Minnesota); Tal S. David, MD (Arthroscopic and Orthopedic Sports Medicine Associates, San Diego, California); Thomas M. DeBerardino, MD (University of Connecticut Health Center, Farmington, Connecticut); Warren R. Dunn, MD, MPH (Vanderbilt University, Nashville, Tennessee); David C. Flanigan, MD (The Ohio State University, Columbus, Ohio); Robert W. Frederick, MD (The Rothman Institute, Philadelphia, Pennsylvania; Thomas Jefferson University, Philadelphia, Pennsylvania); Theodore J. Ganley, MD (Children's Hospital of Philadelphia, Philadelphia, Pennsylvania); Charles J. Gatt Jr, MD (University Orthopedic Associates LLC, New Brunswick, New Jersey); Steven R. Gecha, MD (Princeton Orthopaedic Associates, Princeton, New Jersey); James Robert Giffin, MD (University of Western Ontario, London, Ontario, Canada); Amanda K. Haas, MA (Washington University, St Louis, Missouri); Sharon L. Hame, MD (David Geffen School of Medicine at UCLA, Los Angeles, California); Jo A. Hannafin, MD, PhD (Hospital for Special Surgery, New York, New York); Christopher D. Harner, MD (University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania); Norman Lindsay Harris Jr, MD (Orthopaedic Associates of Aspen & Glenwood, Aspen, Colorado); Keith S. Hechtman, MD (UHZ Sports Medicine Institute, Coral Gables, Florida); Elliott B. Hershman, MD (Lenox Hill Hospital, New York, New York); Rudolf G. Hoellrich, MD (Slocum Research and Education Foundation, Eugene, Oregon); Timothy M. Hosea, MD (University Orthopedic Associates LLC, New Brunswick, New Jersey); Laura J. Huston, MS (Vanderbilt University, Nashville, Tennessee); David C. Johnson, MD, Timothy S. Johnson, MD (National Sports Medicine Institute, Lansdowne, Virginia); Morgan H. Jones, MD (Cleveland Clinic, Cleveland, Ohio); Christopher C. Kaeding, MD (The Ohio State University, Columbus, Ohio); Ganesh V. Kamath, MD (University of North Carolina Medical Center, Chapel Hill, North Carolina); Thomas E. Klootwyk, MD (Methodist Sports Medicine Center–The Orthopedic Specialists, Indianapolis, Indiana); Brett (Brick) A. Lantz, MD (Slocum Research and Education Foundation, Eugene, Oregon); Bruce A. Levy, MD (Mayo Clinic Rochester, Rochester, Minnesota); C. Benjamin Ma, MD (University of California–San Francisco, San Francisco, California); G. Peter Maiers II, MD (Methodist Sports Medicine Center—The Orthopedic Specialists, Indianapolis, Indiana); Barton Mann, PhD (American Orthopaedic Society for Sports Medicine, Chicago, Illinois); Robert G. Marx, MD (Hospital for Special Surgery, New York, New York); Matthew J. Matava, MD (Washington University, St Louis, Missouri); Gregory M. Mathien, MD (Knoxville Orthopedic Clinic, Knoxville, Tennessee); David R. McAllister, MD (David Geffen School of Medicine at UCLA, Los Angeles, California); Eric C. McCarty, MD (University of Colorado Denver School of Medicine, Denver, Colorado); Robert G. McCormack, MD (University of British Columbia, Vancouver, British Columbia, Canada); Bruce S. Miller, MD, MS (University of Michigan, Ann Arbor, Michigan); Carl W. Nissen, MD (Connecticut Children's Medical Center, Hartford, Connecticut); Daniel F. O'Neill, MD, EdD (Littleton Regional Hospital, Littleton, New Hampshire); LTC Brett D. Owens, MD (Keller Army Community Hospital–United States Military Academy, West Point, New York); Richard D. Parker, MD (Cleveland Clinic, Cleveland, Ohio); Mark L. Purnell, MD (Orthopaedic Associates of Aspen & Glenwood, Aspen, Colorado); Arun J. Ramappa, MD (Beth Israel Deaconess Medical Center, Boston, Massachusetts); Michael A. Rauh, MD (State University of New York at Buffalo, Buffalo, New York); Arthur C. Rettig, MD (Methodist Sports Medicine Center–The Orthopedic Specialists, Indianapolis, Indiana); Jon K. Sekiya, MD (University of Michigan, Ann Arbor, Michigan); Kevin G. Shea, MD (Intermountain Orthopedics, Boise, Idaho); Orrin H. Sherman, MD (NYU Hospital for Joint Diseases, New York, New York); James R. Slauterbeck, MD (University of Vermont College of Medicine, Burlington, Vermont); Matthew V. Smith, MD (Washington University, St Louis, Missouri); Jeffrey T. Spang, MD (University of North Carolina Medical Center, Chapel Hill, North Carolina); LTC Steven J. Svoboda, MD (Keller Army Community Hospital–United States Military Academy, West Point, New York); Timothy N. Taft, MD (University of North Carolina Medical Center, Chapel Hill, North Carolina); COL Joachim J. Tenuta, MD (Keller Army Community Hospital–United States Military Academy, West Point, New York); Edwin M. Tingstad, MD (Inland Orthopaedics, Pullman, Washington; Washington State University, Pullman, Washington); Armando F. Vidal, MD (University of Colorado Denver School of Medicine, Denver, Colorado); Darius G. Viskontas, MD (Royal Columbian Hospital, New Westminster, British Columbia, Canada); Richard A. White, MD (University of Missouri–Columbia, Columbia, Missouri); James S. Williams Jr, MD (Cleveland Clinic, Cleveland, Ohio); Michelle L. Wolcott, MD (University of Colorado Denver School of Medicine, Denver, Colorado); Brian R. Wolf, MD (University of Iowa Hospitals and Clinics, Iowa City, Iowa); James J. York, MD (Chesapeake Orthopaedics & Sports Medicine Center, Glen Burnie, Maryland); David C. Harris, BE (Washington University, St Louis, Missouri); Kushal Patel, BA (University of Illinois–Chicago, Chicago, Illinois); David Pearson, BA (Washington University, St Louis, Missouri); Jake Schutzman (Washington University, St Louis, Missouri); Majd Tarabichi (Royal College of Surgeons, Ireland-Bahrain); David Ying, SB (Washington University, St Louis, Missouri).

REFERENCES

1. Allen C, Giffin R, Harner C. Revision anterior cruciate ligament reconstruction. *Orthop Clin N Am*. 2003;34:79-98.
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. Beynnon B, Johnson RJ, Abate JA, Fleming BC, Nichols CE. Treatment of anterior cruciate ligament injuries, part I. *Am J Sports Med*. 2005;33:1579-1602.
4. Brismar B, Wredmark T, Movin T, Leandersson J, Svensson O. Observer reliability in the arthroscopic classification of osteoarthritis of the knee. *J Bone Joint Surg Br*. 2002;84:42-47.
5. Brown C, Carson E. Revision anterior cruciate ligament surgery. *Clin Sports Med*. 1999;18:109-171.
6. Byrt T, Bishop J, Carli JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46:423-429.
7. Cameron M, Briggs K, Steadman J. Reproducibility and reliability of the Outerbridge classification for grading chondral lesions of the knee arthroscopically. *Am J Sports Med*. 2003;31:83-86.

8. Carson E, Anisko E, Restrepo C, Panariello R, O'Brien S, Warren R. Revision anterior cruciate ligament reconstruction. *J Knee Surg*. 2004;17:127-132.

9. Cheatham S, Johnson D. Anatomic revision ACL reconstruction. *Sports Med Arthrosc Rev*. 2010;18:33-39.

10. Cicchetti DV, Feinstein AR. High agreement but low kappa, II: resolving the paradoxes. *J Clin Epidemiol*. 1990;43:551-558.

11. Colvin AC, Shin W, Musahl V, Fu FH. Avoiding pitfalls in anatomic ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2009;17:956-963.

12. Dargel J, Schmidt-Wiethoff R, Fischer S, Mader K, Koebke J, Schneider T. Femoral bone tunnel placement using the transtibial tunnel or the anteromedial portal in ACL reconstruction: a radiographic evaluation. *Knee Surg Sports Traumatol Arthrosc*. 2009;17:220-227.

13. Dunn W, Wolf B, Amendola A, et al. Multirater agreement of arthroscopic meniscal lesions. *Am J Sports Med*. 2004;32:1937-1940.

14. Feinstein AR, Cicchetti DV. High agreement but low kappa, I: the problems of two paradoxes. *J Clin Epidemiol*. 1990;43:543-549.

15. Gavriilidis I, Motsis EK, Pakos EE, et al. Transtibial versus anteromedial portal of the femoral tunnel in ACL reconstruction: a cadaveric study. *Knee*. 2008;15:364-367.

16. George M, Dunn W, Spindler K. Current concepts review: revision anterior cruciate ligament reconstruction. *Am J Sports Med*. 2006;34:2026-2037.

17. Getelman M, Friedman M. Revision anterior cruciate ligament reconstruction surgery. *J Am Acad Orthop Surg*. 1999;7:189-198.

18. Greis P, Johnson D, Fu F. Revision anterior cruciate ligament surgery: causes of graft failure and technical considerations of revision surgery. *Clin Sports Med*. 1993;12:839-852.

19. Grossman M, El Attrache N, Shields C, Glousman R. Revision anterior cruciate ligament reconstruction: three- to nine-year follow-up. *Arthroscopy*. 2005;21:418-423.

20. Harner C, Giffin R, Dunteman R, Annunziata C, Friedman M. Evaluation and treatment of recurrent instability after anterior cruciate ligament reconstruction. *Instr Course Lect*. 2001;50:463-474.

21. Hoehler F. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol*. 2000;53:499-503.

22. Jaureguito J, Paulos L. Why grafts fail. *Clin Orthop Relat Res*. 1996;325:25-41.

23. Javed A, Siddique M, Vaghela M, Hui AC. Interobserver variations in intra-articular evaluation during arthroscopy of the knee. *J Bone Joint Surg Br*. 2002;84:48-49.

24. Johnson DL, Harner CD, Maday MG, Fu FH. Revision anterior cruciate ligament surgery. In: Fu FH, Harner CD, Vince KG, ed. *Knee Surgery*. Baltimore, MD: Williams & Wilkins; 1994:877-895

25. Koh J. Computer-assisted navigation and anterior cruciate ligament reconstruction: accuracy and outcomes. *Orthopedics*. 2005;28(10 suppl):s1283-s1287.

26. Kraemer H. Ramifications of a population model for kappa as a coefficient of reliability. *Psychometrika*. 1979;44:461-472.

27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

28. Lee S, Seong SC, Jo CH, Han HS, An JH, Lee MC. Anterior cruciate ligament reconstruction with use of autologous quadriceps tendon graft. *J Bone Joint Surg Am*. 2007;89(suppl 3):116-126.

29. Marx R, Connor J, Lyman S, et al. Multirater agreement of arthroscopic grading of knee articular cartilage. *Am J Sports Med*. 2005;33:1654-1657.

30. McConkey M, Amendola A, Ramme A, et al. Arthroscopic agreement among surgeons on anterior cruciate ligament tunnel placement. *Am J Sports Med*. 2012;40:2737-2746.

31. Menetrey J, Duthon VB, Laumonier T, Fritschy D. ''Biological failure'' of the anterior cruciate ligament graft. *Knee Surg Sports Traumatol Arthrosc*. 2008;16:224-231.

32. Morgan J, Dahm D, Levy B, Stuart M; MARS Study Group. Femoral tunnel malposition in ACL revision reconstruction. *J Knee Surg*. 2012;25:361-368.

33. Noyes F, Barber-Westin S. Revision anterior cruciate ligament surgery: experience from Cincinnati. *Clin Orthop Relat Res*. 1996;325:116-129.

34. Sommer C, Friederich N, Müller W. Improperly placed anterior cruciate ligament grafts: correlation between radiological parameters and clinical results. *Knee Surg Sports Traumatol Arthrosc*. 2000;8:207-213.

35. Spindler K. The Multicenter ACL revision study (MARS): a prospective longitudinal cohort to define outcomes and independent predictors of outcomes for revision anterior cruciate ligament reconstruction. *J Knee Surg*. 2007;20:303-307.

36. Warme B, Ramme A, Willey M, et al; MOON Knee Group. Reliability of early postoperative radiographic assessment of tunnel placement after anterior cruciate ligament reconstruction. *Arthroscopy*. 2012;28:942-951.

37. Wolf B, Ramme A, Britton C, Amendola A; MOON Knee Group. Anterior cruciate ligament tunnel placement. *J Knee Surg*. 2014;27:309-318.

38. Wolf B, Ramme A, Wright R, et al; MOON Knee Group. Variability in ACL tunnel placement: observational clinical study of surgeon ACL tunnel variability. *Am J Sports Med*. 2013;41:1265-1273.

39. Wright J. A practical guide to assigning levels of evidence. *J Bone Joint Surg Am*. 2007;89:1128-1130.