Multi-rater Agreement in the Assessment of Anterior Cruciate Ligament Reconstruction Failure. A Radiographic and Video Analysis of the MARS Cohort

Matthew J. Matava, MD,
Washington University, St. Louis

Robert A. Arciero, MD,
University of Connecticut Health Center

Keith M. Baumgarten, MD,
Orthopedic Institute

James L. Carey, MD, MPH,
University of Pennsylvania

Thomas M. DeBerardino, MD,
University of Connecticut Health Center

Sharon L. Hame, MD,
David Geffen School of Medicine at UCLA

Jo A. Hannafin, MD, PhD,
Hospital for Special Surgery

Bruce S. Miller, MD, MS,
University of Michigan

Carl W. Nissen, MD,
Connecticut Children's Medical Center

Timothy N. Taft, MD,
University of North Carolina Medical Center

Brian R. Wolf, MD,
University of Iowa Hospitals and Clinics

Rick W. Wright, MD, and
Washington University, St. Louis

MARS Group

Abstract

Background—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. There is limited data regarding the agreement among orthopedic surgeons in terms of the etiology of primary ACL reconstruction failure and accuracy of graft tunnel placement.

Robert W. Frederick, MD — The Rothman Institute/Thomas Jefferson University

Theodore J. Ganley, MD — Children's Hospital of Philadelphia

Elizabeth A. Garofoli — Washington University, St. Louis

Charles J. Gatt, Jr., MD — University Orthopedic Associates, LLC

Steven R. Gecha, MD — Princeton Orthopaedic Associates

James Robert Giffin, MD — Fowler Kennedy Sports Medicine Clinic- University of Western Ontario

Amanda K. Haas, MA — Washington University, St. Louis

Christopher D. Harner, MD — University of Pittsburgh Medical Center

Timothy M. Hosea, MD — University Orthopedic Associates, LLC

Laura J. Huston, MS — Vanderbilt University

David C. Johnson, MD — National Sports Medicine Institute

Timothy S. Johnson, MD — National Sports Medicine Institute

Morgan H. Jones, MD — Cleveland Clinic

Christopher C. Kaeding, MD — The Ohio State University

Ganesh V. Kamath, MD — University of North Carolina Medical Center

Thomas E. Klootwyk, MD — Methodist Sports Medicine Center-The Orthopedic Specialists

Brett (Brick) A. Lantz, MD — Slocum Research and Education Foundation

Bruce A. Levy, MD — Mayo Clinic Rochester

C. Benjamin Ma, MD — University of California, San Francisco

G. Peter Maiers, II, MD — Methodist Sports Medicine Center-The Orthopedic Specialists

Robert G. Marx, MD — Hospital for Special Surgery

Gregory M. Mathien, MD — Knoxville Orthopedic Clinic

David R. McAllister, MD — David Geffen School of Medicine at UCLA

Eric C. McCarty, MD — University of Colorado Denver School of Medicine

Robert G. McCormack, MD — University of British Columbia

Sam K. Nwosu, MS — Vanderbilt University

Daniel F. O'Neill, MD, Ed.D — Littleton Regional Hospital

LTC Brett D. Owens, MD — Keller Army Community Hospital-United States Military Academy

Richard D. Parker, MD — Cleveland Clinic

Mark L. Purnell, MD — Orthopaedic Associates of Aspen & Glenwood

Arun J. Ramappa, MD — Beth Israel Deaconess Medical Center

Michael A. Rauh, MD — State University of New York at Buffalo

Arthur C. Rettig, MD — Methodist Sports Medicine Center-The Orthopedic Specialists

Jon K. Sekiya, MD — University of Michigan

Kevin G. Shea, MD — Intermountain Orthopedics

Orrin H. Sherman, MD — NYU Hospital for Joint Diseases

James R. Slauterbeck, MD — University of Vermont College of Medicine

Matthew V. Smith, MD — Washington University, St. Louis

Jeffrey T. Spang, MD — University of North Carolina Medical Center

Kurt P. Spindler, MD — Vanderbilt University

Michael J. Stuart, MD — Mayo Clinic Rochester

LTC. Steven J. Svoboda, MD — Keller Army Community Hospital-United States Military Academy

COL Joachim J. Tenuta, MD — Keller Army Community Hospital-United States Military Academy

Edwin M. Tingstad, MD — Inland Orthopaedics/Washington State University

Armando F. Vidal, MD — University of Colorado Denver School of Medicine

Darius G. Viskontas, MD — Royal Columbian Hospital

Richard A. White, MD — University of Missouri-Columbia

James S. Williams Jr, MD — Cleveland Clinic

Michelle L. Wolcott, MD — University of Colorado Denver School of Medicine

James J. York, MD — Chesapeake Orthopaedics & Sports Medicine Center

Purpose—The purpose of this study is to test the hypothesis that experienced knee surgeons

have a high level of inter-observer reliability in the agreement of the etiology of the primary ACL reconstruction failure, anatomical graft characteristics, tunnel placement.

Methods—Twenty cases of revision ACL reconstruction were randomly selected from the 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. 10 MARS surgeons not involved with the primary surgery reviewed all 20 cases. Each surgeon completed a two-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 etiology of graft failure. Inter-rater agreement was determined for each question. Inter-rater agreement was determined for each question with the kappa coefficient and prevalence adjusted bias adjusted kappa (PABAK).

Results—The 10 reviewers were in practice an average of 14 years. All 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 etiology of ACL graft failure. When specifically asked about technical error as the cause for failure, inter-observer agreement was only slight (prevalence adjusted bias adjusted kappa [PABAK]: 0.26). There was fair overall agreement on ideal femoral tunnel placement (PABAK: 0.55), but only slight agreement whether a femoral tunnel was too anterior (PABAK: 0.24) and fair agreement 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 etiology of primary ACL graft failure as well as the ideal femoral and tibial tunnel placement in patients undergoing revision ACL reconstruction.

Introduction

Anterior cruciate ligament (ACL) tears are a common cause of disability to 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 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[15,20,32,33,35].

Understanding the specific etiology of primary reconstruction failure is paramount to improving revision ACL reconstruction. It has been widely assumed, based on Level 4 and 5[39] case series and expert opinion[9,17,18,23], that technical errors[1,12,16,18,22,24,31,34] – in particular femoral tunnel malposition[31,34,37] – 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 in order to better evaluate those factors potentially influencing patient outcome. However, for this multicenter format to be effective, it is imperative that participating surgeons provide reliable and reproducible clinical data, agreeing upon the anatomical 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 will 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 inter-observer agreement among experienced sports medicine specialists participating in the MARS group regarding specific technical and anatomic factors of the primary graft as well as the etiology of the primary reconstruction failure. Assessments were made based on the patient history, standardized radiographs, and arthroscopic videos of the failed index surgery. The results of this multi-center trial could be instrumental in determining the factor(s) 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 comprised 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 six-hour training course outlining the goals of the MARS study 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 five years of clinical experience, describing the purposes of the study and to determine whether or not 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 co-morbidities. 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 three 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, etiology of primary ACL tear, date of primary ACL reconstruction, primary graft source, method of fixation, prior surgical technique if known (e.g. one-incision trans-tibial, one-incision anteromedial portal, two-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 30-second (approximate) segment of the arthroscopic video taken during the revision surgery using a 30° arthroscope placed in the anterolateral and anteromedial portals. Each video demonstrated the surgeon probing the failed graft (if present) in order to assess graft attenuation or absence. The femoral and tibial tunnel apertures and their size once the failed graft had been debrided and hardware removed by the operating surgeon was also shown.

A compilation of all 20 case histories, corresponding radiographs, and arthroscopic videos were 'burned' to a digital video disc (DVD) and sent to a random selection of 10 participating MARS surgeons possessing at least five years of clinical experience. Each surgeon was asked to complete a two-part questionnaire regarding each case (Appendix). Part I was comprised of six questions that dealt solely with each surgeon's practice type and experience in performing primary and revision ACL surgery. Part II was comprised of 21 questions that required a simple 'yes' or 'no' answer that 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 (i.e. absent, present but elongated, torn), placement of the femoral and tibial tunnels (i.e. too anterior, too posterior, too vertical, etc.), and the surgeon's opinion as to the cause of failure (i.e. traumatic, biologic, technical, combination). The reviewers were not given any sort of primer or instruction on the objective 'gold standard' or predetermined 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

Inter-rater agreement regarding the various responses on Part II of the questionnaire was analyzed by using several different measures. Inter-rater agreement (with confidential intervals [C.I.]) was determined for each question by calculating the percent perfect agreement among all pair-wise, between-rater comparisons. There were 56 pair-wise comparisons. Those comparisons where 8 of 10 surgeons agreed on the question of interest were also determined since this number (80%) was felt to represent reasonably high degree of agreement. Statistical analyses were performed using SAS software, version 9.2 (SAS Institute Inc., Cary, NC, USA). This provided a percentage of agreement between raters. In addition the number of cases in which 8/10 reviewers agreed was also determined for each question.

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

$$K = \frac{\text{Observed agreement} - \text{Chance agreement}}{1 - \text{Chance agreement}}$$

Interpreting K using this model, however, assumes a generally equal distribution of prevalence of the studied attribute. If prevalence is not equally distributed then the K value is distorted and becomes less meaningful. The Prevalence Index (P.I.) was used to determine the appropriateness of the K value and was calculated with the following equation:

$$\text{P.I.} = \frac{[\#\text{of Yes} - \text{Yes Responses} - \#\text{of No} - \text{No Responses}]}{\text{Total} \# \text{of Responses}}$$

The P.I. ranges from 0 to 1, with a higher P.I. indicating that K is less likely to accurately evaluate agreement due to the problems of an uneven prevalence distribution. A prevalence adjusted bias adjusted K (PABAK) is one method for adjusting K for the paradoxes caused by large differences between the two types of agreement (prevalence) or the two types of disagreement (bias)[6]. A PABAK is particularly useful in cases with high percentage agreement but a low K 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 K with values below 0.0 suggesting poor agreement, a K value of 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 agreement, and 0.81 to 1.00: almost perfect agreement. This classification was originally developed in the study of agreement between two raters, where the K coefficient reflects error, not low prevalence. This ordinal interpretation scheme was used for the PABAK scores as a descriptor of inter-observer agreement. Their 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 to the PABAK to avoid misleading interpretations that reflect the nature of the population rather than the observation procedure itself.[26]

Results

The 10 reviewers were in practice an average of 14 years (range, 5 to 35 years). All performed at least 25 ACL reconstructions per year and 9 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 performed annually by the reviewers was 1 to 5 revisions: 4, 6 to 10 revisions: 3, 11 to 15 revisions: 1, 16 to 20 revisions: 0, and greater than 20 revisions: 2.

Overall inter-observer agreement (with agreement greater than 80%) is shown in Table 1. The K, P.I., and PABAK values were calculated for each question and shown in Table 2. Topics with P.I. values closer to 1.0 indicate a decreased relevance of the K values and reflect the increased influence of prevalence resulting in larger discrepancies between their K and PABAK values.

Inter-observer agreement for questions regarding the failed graft's presence and condition averaged 87% (range, 83% to 90%). At least 80% of the reviewers agreed in 90% (18/20) of the cases. The percent agreement among the reviewers regarding the specific etiology of graft failure averaged 72% (range, 52% to 98%). At least 80% of the reviewers agreed on the etiology of graft failure in only 55% of the cases. Inter-observer agreement was highest regarding whether or not other ligamentous insufficiency was the primary cause (98%) (PABAK: 0.96) and lowest if a combination of factors was the likely etiology of failure (52%) (PABAK: 0.04). At least 80% of the reviewers agreed in 55% (11/20) of cases when estimating the etiology of failure.

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

Inter-observer agreement was 77% (95% C.I.: 75% to 80%) (PABAK: 0.55) when determining if the femoral tunnel was ideal in placement compared to 58% (95% C.I.: 55% to 62%) (PABAK: 0.17) agreement 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 the percent agreement for questions regarding specific femoral tunnel placement (i.e. too anterior, too posterior, too vertical) averaged 76% (range, 62% to 92%). At least 80% of the reviewers agreed in 65% (13/20). Inter-observer agreement was highest when evaluating posterior femoral tunnel placement (92%) (PABAK: 0.84) and lowest when assessing anterior placement (62%) (PABAK: 0.24). Agreement regarding femoral tunnel verticality averaged 73% (range, 70% to 76%) (PABAK: 0.46).

The percent agreement for questions about specific tibial tunnel placement averaged 84% (range, 72% to 90%) with at least 80% of the reviewers agreeing on specific tibial tunnel position in 83% (17/20) of the cases. Inter-observer agreement was highest when evaluating lateral tibial tunnel placement (PABAK: 0.81) and lowest when assessing posterior placement (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, there was significant variability in agreement regarding the surmised etiology of graft failure

and tunnel placement. Overall inter-observer agreement was 77% (PABAK: 0.55) when determining if the femoral tunnel was ideal in placement and size compared to 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 agreement with the work of Morgan et al.[31] 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 to the tibial tunnel.

Prior research has attempted to determine the inter-observer reliability in the assessment of femoral tunnel placement in primary ACL reconstruction. Warme et al.[37] prospectively evaluated the postoperative plain radiographs of 54 patients following primary ACL reconstruction. Three blinded reviewers performed eight different radiographic measurements to assess tunnel location. Intra-observer reliability for femoral measurements ranged from none to substantial, but was moderate to almost perfect for tibial tunnel measurements. Inter-observer 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 following ACL reconstruction are quite variable, with reliability falling only into the fair to moderate categories.

Wolf et al.[36] analyzed variation in ACL tunnel placement between surgeons and the influence of preferred surgical technique and surgeon experience using three-dimensional computed tomography. There was a relatively high degree of intra-surgeon 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. They found that operating surgeons were more likely to judge their own tunnels more favorably than other observers. However, independent surgeon reviewers appeared to be more critical of other surgeons' tunnels. They concluded that, overall, surgeons do not agree on the ideal placement for single-bundle ACL tunnels.

Multiple studies have assessed the inter-observer and intra-observer reliability in the arthroscopic evaluation and classification of other knee pathology.[2,4,7,13,29] Marx et al.[29] demonstrated good inter-observer reliability with arthroscopic classification of articular

cartilage lesions. Six experienced surgeons based on video analysis classified thirty-one different lesions. The authors reported 81% to 94% agreement depending on lesion location; however, K values varied between fair to near perfect agreement (range: 0.34 to 0.87).

Brismar et al.[4] studied both intra-observer and inter-observer 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. They found 59% to 62% overall inter-observer agreement, and 55% to 77% intra-observer agreement with K values indicating moderate agreement. Studies by both Anderson et al.[2] and Dunn et al.[13] used intra-operative 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 the impact of the prevalence paradox in their study, with certain categories having low K values despite high observed agreement. They specifically noted that their conclusions were based on their percent agreement rather than K coefficients due to the problems related to K.

While Cohen's K coefficient has been used in multiple reliability studies to assess agreement, K had limited usefulness in this study due to several well-documented problems. These problems, referred to as "paradoxes", limit K's application and interpretation leading many statisticians to warn against using K alone to evaluate agreement. The paradox of prevalence was particularly relevant to this study and can be responsible for significantly depressed K 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 to the converse (i.e. '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 inter-observer 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), K was significantly decreased and actually resulted in a negative value (-0.0152). The high P.I. value (0.96), however, forewarns of significant K distortion. Similar problems due to the prevalence paradox were seen throughout this study. One method of resolving this dilemma is adjusting the K coefficient by using the mean of the observed agreement and disagreement when calculating the chance agreement factor. This adjustment, referred to as PABAK, eliminates this problem (the K paradox) caused by uneven distribution of prevalence and bias[6,10,14]. PABAK values, shown in Figure 2, reflect agreement without the influence of prevalence or bias. It should be noted that similar to K, statisticians have warned against using PABAK alone to interpret agreement as prevalence and bias do have informative value in assessing agreement[10,26]. We used the Landis and Koch classification to provide useful benchmarks to interpret agreement. This classification was developed in the study of agreement between two raters, where the K coefficient reflects error, not low prevalence. Due to the low prevalence found in our study and the associated impact upon kappa, we have reported multiple statistics (i.e., kappa, prevalence index, and PABAK) so that each reader can interpret the adequacy of agreement within their specific context.

There are several limitations to this study that should be addressed. The use of video, while consistently used for reliability studies,[2,4,13,29] does have 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 we would suggest 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 (i.e. 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 prior to 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 cause(s) of graft failure and the accuracy of graft placement. Finally, the results of this study are potentially limited by the fact that experienced orthopedic 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 following 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 utilized 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 to prior literature dealing with ACL revision. This is the first study, to our knowledge, to assess the inter-observer reliability when evaluating primary ACL reconstruction failure, particularly focusing on graft location, anatomical graft characteristics, and etiology 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 in order 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 in order 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 may have obtained with less experienced surgeons, we feel justified in using a more experienced group since most ACL revisions are theoretically performed by more experienced knee surgeons.

In conclusion, there was wide variability in agreement among knee experts as to the specific etiology of primary ACL reconstruction failure and the appropriateness of tunnel placement in patients undergoing ACL revision. Inter-observer agreement was only slight when attributing the cause of primary graft failure 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 whether a femoral tunnel was too anterior and fair agreement 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 etiology of primary ACL graft failure as well as ideal tunnel location in order to improve the outcome of ACL reconstruction as well as to facilitate future research in revision ACL surgery.

Acknowledgments

Appendix

MARS Video Study Reviewer Form

Part I

1. Are you fellowship-trained in sports medicine?

    a. Yes

    b. No

2. Number of years you have been in practice:

    a. 0-5 years

    b. 6-10 years

    c. 11-15 years

    d. 16-20 years

    e. > 20 years

3. How would you describe your practice?

    a. Private with no academic affiliation Private with a clinical affiliation with an academic center

      b. Full-time academic

4. What percentage of your practice is related to surgery of the knee?

      a. 1% –25%

      b. 26% –50%

      c. 51% –75%

      d. 75% –100%

5. Number of PRIMARY ACL reconstructions you perform per year:

      a. 1-25

      b. 26-50

      c. 51-75

      d. 76-100

      e. > 100

6. Number of REVISION ACL reconstructions you perform per year:

      a. 1-5

      b. 6-10

      c. 11-15

      d. 16-20

      e. > 20

Part II

For the following cases please use the corresponding patient clinical history, radiographs, and surgical videos to formulate your answers. Please select only <u>one</u> answer for each question.

1. In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel ideal in terms of both position AND size?

      a. Yes

      b. No

2. Do you feel that PRIMARY GRAFT FAILURE was due to insufficient FEMORAL FIXATION?

      a. Yes

      b. No

3. In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel ideal in terms of position, but ENLARGED?

      a. Yes

b. No

4. In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO VERTICAL?

    a. Yes

    b. No

5. In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO ANTERIOR?

    a. Yes

    b. No

6. In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO POSTERIOR?

    a. Yes

    b. No

7. Do you feel that PRIMARY GRAFT FAILURE was due to insufficient TIBIAL FIXATION?

    a. Yes

    b. No

8. In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of both position AND size?

    a. Yes

    b. No

9. In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of position, but ENLARGED?

    a. Yes

    b. No

10. In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO MEDIAL?

    a. Yes

    b. No

11. In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO LATERAL?

    a. Yes

    b. No

12. In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO ANTERIOR?

    a. Yes

    b. No

13. In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO POSTERIOR?

    a. Yes

    b. No

14. In terms of the APPEARANCE of the failed ACL graft, is the graft ABSENT?

    a. Yes

    b. No

15. In terms of the APPEARANCE of the failed ACL graft, is the graft PRESENT, but ELONGATED?

    a. Yes

    b. No

16. In terms of the APPEARANCE of the failed ACL graft, is the graft PRESENT, but the MAJORITY TORN?

    a. Yes

    b. No

17. In terms of the ETIOLOGY of primary graft failure, do you feel the cause was TRAUMATIC?

    a. Yes

    b. No

18. In terms of the ETIOLOGY of primary graft failure, do you feel the cause was BIOLOGIC FAILURE TO HEAL?

    a. Yes

    b. No

19. In terms of the ETIOLOGY of primary graft failure, do you feel the cause was TECHNICAL ERROR from the prior surgery?

    a. Yes

    b. No

20. Is terms of the ETIOLOGY of primary graft failure, do you feel the cause was due to a COMBINATION of factors (i.e. traumatic, biologic, and/or technical)?

    a. Yes

    b. No

21. Is there evidence from the available data that primary graft failure was due to OTHER ligamentous insufficiency (i.e. lateral collateral, posterolateral corner, medial collateral)?

    a. Yes

    b. No

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, Beaufils P, Cohen M, Cole BJ, Coolican M, Ferretti M, Glenn RE Jr, Johnson R, Neyret P, Ochi M, Panarella L, Siebold R, Spindler KP, Ait Si Selmi T, Verdonk P, Verdonk R, Yasuda K, Kowalchuk DA. Interobserver reliability of the International Society of Arthroscopy, Knee Surgery and Orthopaedic Sports Medicine (ISAKOS) classification of meniscal tears. Am J Sports Med. 39(5):926–32. [PubMed: 21411745]

3. Beynnon B, et al. Treatment of anterior cruciate ligament injuries, part I. Am J Sports Med. 2005; 33:1579–602. [PubMed: 16199611]

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. [PubMed: 11837831]

5. Brown C, Carson E. Revision anterior cruciate ligament surgery. Clin Sports Med. 1999; 18:109–171. [PubMed: 10028120]

6. Byrt T, Bishop J, Carli JB. Bias, prevalence and kappa. J Clin Epidemiol. 1993; 46:423–429. [PubMed: 8501467]

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. [PubMed: 12531763]

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. [PubMed: 15366266]

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. [PubMed: 2189948]

11. Colvin A, et al. Avoiding pitfalls in anatomic ACL reconstruction. Knee Surg, Sports Traumatol, Arthrosc. 2009; 17:956–63. [PubMed: 19399476]

12. Dargel J, et al. 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. [PubMed: 18843479]

13. Dunn W, Wolf B, Amendola A, et al. Multirater agreement of arthroscopic meniscal lesions. Am J Sports Med. 2004; 32:1937–1940. [PubMed: 15572324]

14. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990; 43:543–549. [PubMed: 2348207]

15. Gavriilidis I, et al. Transtibial versus anteromedial portal of the femoral tunnel in ACL reconstruction: a cadaveric study. Knee. 2008; 15:364–367. [PubMed: 18583137]

16. George M, Dunn W, Spindler K. Current concepts review: revision anterior cruciate ligament reconstruction. Am J Sports Med. 2006; 34:2026–2037. [PubMed: 17092921]

17. Getelman M, Friedman M. Revision anterior cruciate ligament reconstruction surgery. J Am Acad Orthop Surg. 1999; 7:189–98. [PubMed: 10346827]

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; 12:839–52. [PubMed: 8261529]

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. [PubMed: 15800521]

20. Harner C, Giffin R, Dunteman R, Annunziata C, Friedman M. Evaluation and treatment of recurrent instability after anterior cruciate ligament reconstruction. AAOS Instructional Course Lectures. 2001; 50:463–474. chap. 47.

21. Hoehler F. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. J Clin Epidemiol. 2000; 53:499–503. [PubMed: 10812322]

22. 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-B:48–49. [PubMed: 11838444]

23. Jaureguito J, Paulos L. Why grafts fail. Clin Orthop Relat Res. 1996; 325:25–41. [PubMed: 8998884]

24. Johnson, D., et al. Revision anterior cruciate ligament surgery. In: Fu, FH.; Harner, CD.; Vince, KG., editors. Knee Surgery. Williams & Wilkins; Baltimore: 1994.

25. Koh J. Computer-assisted navigation and anterior cruciate ligament reconstruction: accuracy and outcomes. Orthopedics. 2005; 28(10 Suppl):s1283–7. [PubMed: 16235456]

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. [PubMed: 843571]

28. Lee S, et al. Anterior cruciate ligament reconstruction with use of autologous quadriceps tendon graft. J Bone Joint Surg Am. 2007; 89(Suppl 3):116–26. [PubMed: 17908877]

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. [PubMed: 16093545]

30. McConkey M, Amendola A, Ramme A, Dunn W, Flanigan D, Britton C, MOON Knee Group. Wolf B, Spindler K, Carey J, Cox C, Kaeding C, Wright R, Matava M, Brophy R, Smith M, McCarty E, Vida A, Wolcott M, Marx R, Parker R, Andrish J, Jones M. Arthroscopic agreement among surgeons on anterior cruciate ligament tunnel placement. Am J Sports Med. 2012; 40:2737–2746. [PubMed: 23075806]

31. 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. [PubMed: 23150344]

32. Menetrey J, et al. Biological failure" of the anterior cruciate ligament graft. Knee Surg Sports Traumatol Arthrosc. 2008; 16:224–31. [PubMed: 18183368]

33. Noyes F, Barber-Westin S. Revision anterior cruciate ligament surgery: experience from Cincinnati. Clin Orthop Relat Res. 1996; 325:116–29. [PubMed: 8998864]

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. [PubMed: 10975260]

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. [PubMed: 17993074]

36. Wolf B, Ramme A, Britton C, Amendola A. MOON Knee Group. Anterior cruciate ligament tunnel placement. J Knee Surg. 2014; 27:309–18. [PubMed: 24414390]

37. Warme B, Ramme A, Willey M, Britton C, Flint J, Amendola A, Wolf B. MOON Knee Group. Reliability of early postoperative radiographic assessment of tunnel placement after anterior cruciate ligament reconstruction. Arthroscopy. 2012; 28:942–51. [PubMed: 22381687]

38. Wolf B, Ramme A, Wright R, Brophy R, McCarty E, Vidal A, Parker R, Andrish JT, Amendola A. MOON Knee Group. Variability in ACL tunnel placement: observational clinical study of surgeon ACL tunnel variability. Am J Sports Med. 2013; 41:1265–73. [PubMed: 23618702]

39. Wright J. A practical gide to assigning levels of evidence. J Bone Joint Surg Am. 2007; 89:1128–1130. [PubMed: 17473152]

What is known about this subject?

There has been increased interest in intra-observer agreement as to the proper placement of both the femoral and tibial tunnels for ACL reconstruction. Thus far, there has been only fair agreement among surgeons as to where the tunnels should be placed. It is not known to what degree knee experts agree on appropriately placed ACL grafts or the actual etiology of failure in those patients undergoing revision ACL reconstruction.

What this study adds to existing knowledge

This study shows that there is wide variability among knee experts as to the theorized cause of ACL reconstruction failure. In addition, the agreement regarding the appropriate placement of the femoral and tibial tunnels was only fair. Therefore, more objective criteria are needed to accurately determine the etiology of primary ACL graft failure as well as the ideal femoral and tibial tunnel placement in patients undergoing revision ACL reconstruction.
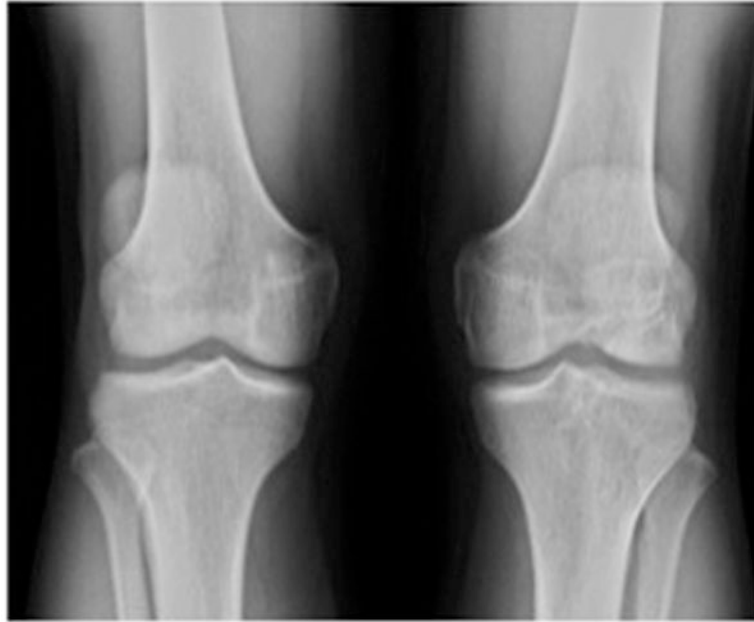
Figure 1.
Case #1. Selected radiographic views: 1a: Weight bearing anteroposterior; 1b: 45° flexion-weight bearing (Rosenberg); 1c: 30° lateral; 1d: Full extension lateral.

Selected questions pertaining to tunnel location and number of corresponding "Yes" or "No" responses:

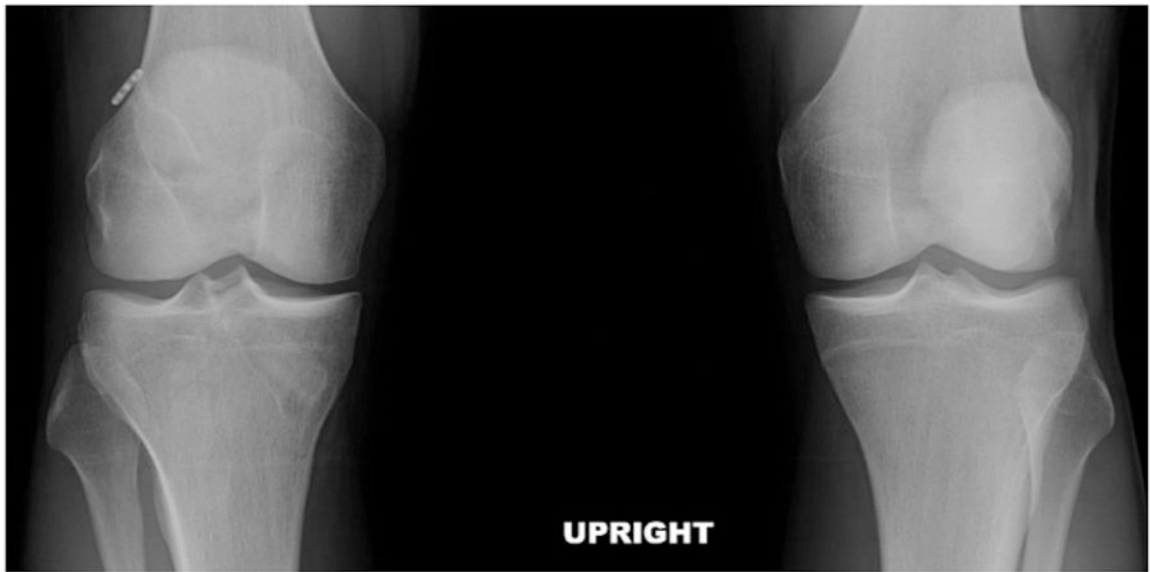| Question	Yes	No
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel ideal in terms of both position AND size?	Yes: 0	No: 10
Do you feel that PRIMARY GRAFT FAILURE was due to insufficient FEMORAL FIXATION?	Yes: 0	No: 10
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel ideal in terms of position, but ENLARGED?	Yes: 0	No: 10
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO VERTICAL?	Yes: 3	No: 7
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO ANTERIOR?	Yes: 9	No: 1
In regard to the PRIOR FEMORAL tunnel position at revision, is the tunnel TOO POSTERIOR?	Yes: 0	No: 10
Do you feel that PRIMARY GRAFT FAILURE was due to insufficient TIBIAL FIXATION?	Yes: 0	No: 10
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of both position AND size?	Yes: 1	No: 9
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of position, but ENLARGED?	Yes: 1	No: 9
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO MEDIAL?	Yes: 0	No: 10
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO LATERAL?	Yes: 2	No: 8
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO ANTERIOR?	Yes: 8	No: 2
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO POSTERIOR?	Yes: 0	No: 10

Figure 2.
Case #2. Selected radiographic views: 2a: Weight bearing anteroposterior; 2b: 45° flexion-weight bearing (Rosenberg); 2c: 30° lateral; 2d: Full extension lateral.

Selected questions pertaining to tunnel location and number of corresponding "Yes" or "No" responses:

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

In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of both position AND size?	Yes: 4	No: 6
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel ideal in terms of position, but ENLARGED?	Yes: 4	No: 6
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO MEDIAL?	Yes: 0	No: 10
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO LATERAL?	Yes: 1	No: 9
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO ANTERIOR?	Yes: 0	No: 10
In regard to the PRIOR TIBIAL tunnel position at revision, is the tunnel TOO POSTERIOR?	Yes: 3	No: 7

Table 1

Inter-Observer Agreement For Each Radiographic And Graft-Related Topic Assessed

	Overall Agreement (%)	95% Confidence Interval	# of Cases with >80% agreement
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
Etiology 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

Table 2

The Kappa Coefficient, Prevalence Index, and PABAK Values for Each Radiographic and Graft-related Topic Assessed

	Kappa Coefficient	Prevalence Index	PABAK
Ideal Placement			
Femur	0.16	0.68	0.55
Tibia	0.15	0.13	0.17
Femoral Positioning			
Vertical	0.37	0.40	0.46
Anterior	0.24	0.04	0.24
Posterior	0.18	0.9	0.84
Tibial Positioning			
Medial	0.37	0.81	0.78
Lateral	0.16	0.88	0.81
Anterior	0.32	0.74	0.69
Posterior	0.19	0.55	0.43
Tunnel Size			
Femur	0.20	0.85	0.78
Tibia	0.14	0.63	0.48
Graft Condition			
Absent	0.81	0.066	0.82
Elongated	0.70	0.28	0.72
Torn	0.45	0.63	0.67
Graft Fixation			
Femur	0.40	0.86	0.84
Tibia	-0.015	0.96	0.92
Etiology of Failure			
Trauma	0.43	0.058	0.40
Biologic	0.095	0.70	0.53
Technical	0.056	0.47	0.26
Combination	0.056	0.015	0.04
Other ligament injury	0.10	0.98	0.96

PABAK: Prevalence Adjusted Bias Adjusted Kappa