The Statistical Fragility of Single-Bundle vs Double-Bundle Autografts for ACL Reconstruction

A Systematic Review of Comparative Studies

Cooper B. Ehlers,*[†] MD, MS, Andrew J. Curley,[‡] MD, Nathan P. Fackler,[†] MS,
Arjun Minhas,[†] MD, MS, Ariel N. Rodriguez,[†] MS, Kory Pasko,[†] BS, and Edward S. Chang,[§] MD
Investigation performed at MedStar Georgetown University Hospital, Washington, DC, USA

Background: The statistical significance of a given study outcome can be liable to small changes in findings. *P* values are common, but imperfect statistical methods to convey significance, and inclusion of the fragility index (FI) and fragility quotient (FQ) may provide a clearer perception of statistical strength.

Purpose/Hypothesis: The purpose was to examine the statistical stability of studies comparing primary single-bundle to double-bundle anterior cruciate ligament reconstruction (ACLR) utilizing autograft and independent tunnel drilling. It was hypothesized that the study findings would be vulnerable to a small number of outcome event reversals, often less than the number of patients lost to follow-up.

Study Design: Systematic review; Level of evidence, 2.

Methods: Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, the authors searched PubMed for comparative studies and randomized controlled trials (RCTs) published in select journals, based on impact factor, between 2005 and 2020. Risk-of-bias assessment and methodology scoring were conducted for the included studies. A total of 48 dichotomous outcome measures were examined for possible event reversals. The FI for each outcome was determined by the number of event reversals necessary to alter significance. The FQ was calculated by dividing the FI by the respective sample size.

Results: Of the 1794 studies screened, 15 comparative studies were included for analysis; 13 studies were RCTs. Overall, the mean FI and FQ were 3.14 (IQR, 2-4) and 0.050 (IQR, 0.032-0.062), respectively. For 72.9% of outcomes, the FI was less than the number of patients lost to follow-up.

Conclusion: Studies comparing single-bundle versus double-bundle ACLR may not be as statistically stable as previously thought. Comparative studies and RCTs are at substantial risk for statistical fragility, with few event reversals required to alter significance. The reversal of fewer than 4 outcome events in a treatment group can alter the statistical significance of a given result; this is commonly less than the number of patients lost to follow-up. Future comparative study analyses might consider including FI and FQ with *P* values in their statistical analysis.

Keywords: fragility index; fragility quotient; statistical significance; *P* value; anterior cruciate ligament reconstruction; autograft

Injuries to the anterior cruciate ligament (ACL) are common in athletes, with more than 2 million occurring worldwide annually.[38] Surgical management of this ligament through ACL reconstruction (ALCR) aims to restore knee function, stability, and preinjury levels of activity.[33] The techniques for femoral and tibial tunnel placement and graft type selection vary among surgeons,[39] but a recent systematic review[16] found that single-bundle reconstruction with independent tunnel drilling seemed to be the current preferred technique in the United States. These evidence-based decisions are driven by the findings from randomized controlled trials (RCTs), which support the highest-level recommendations produced by the American Academy of Orthopaedic Surgeons (AAOS).[9] However, the statistical stability of these studies may be more fragile than previously thought.

The importance of data from comparative studies and RCTs is commonly conveyed via various test statistics and

statistical thresholds. One common test statistic is the *P* value, which is compared with an arbitrarily chosen α threshold, typically set at α = 0.05. If the *P* value is less than this threshold, the null hypothesis is rejected. Thus, there is a less than 5% chance that the collected data occurred due to random chance,[6] and it is generally accepted that this difference is statistically significant. However, implementation of the *P* value is imperfect. It has been shown to be malleable to study design, randomization, and study power.[18] Importantly, significance can be altered by a small number of event reversals within a sample,[17,23,30,44] and the intent of amalgamated data analyzed by systematic reviews as well as meta-analyses is to mitigate the statistical lability of any single study. If the number of event reversals required to alter significance is less than the number of patients lost to follow-up,[45] it could be possible for studies to have altered findings simply by maintaining follow-up. The fragility index (FI) is a numerical representation of statistical robustness versus fragility.[12] FI is calculated as the number of outcome event reversals necessary to convert a finding from significant to nonsignificant or vice versa. The inclusion of this statistic can enhance the information portrayed by *P* values, but it exists independent of sample size and is similarly limited. To address these issues, the fragility quotient (FQ) was introduced. The FQ is determined by dividing the FI by the sample size.[3] For a given outcome, the FQ represents the percentage of reversals required to alter statistical significance. Larger values for FI and FQ imply stronger statistical stability, whereas lower values suggest statistical fragility. Despite the ability to provide further statistical insight for clinicians, fragility is not generally reported in RCTs and comparative studies.

To our knowledge, there have been no studies applying fragility analysis to comparative studies and RCTs regarding the different graft bundle options for ACLR. The purpose of this study was to determine the statistical stability of studies comparing single-bundle and double-bundle autografts in primary ACLR with independent tunnel drilling. The primary objective for this study was to calculate the mean FI and mean FQ for dichotomous outcomes reported by these studies. The secondary aim for this study was to perform subgroup analysis and calculate the proportion of outcome events for which FI is less than the number of patients lost to follow-up. We hypothesize that the findings of these studies are vulnerable to a small number of outcome event reversals and that the number of outcome event reversals are often less than the number of patients lost to follow-up.

METHODS

A systematic review was performed according to *Cochrane Handbook for Systematic Reviews*. No approval by the institutional review board was necessary for this story. The study search overview is described in Figure 1.

Search Strategy

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Relevant literature searches were performed via PubMed for select journals. The articles must have included comparative studies and RCTs pertaining to the utilization of single-bundle and double-bundle autografts published in select journals from 2005 to 2020. The select journals were chosen for their prominence within the field of orthopaedic surgery and sports medicine. The 10 orthopaedic journals included were the *British Journal of Sports Medicine*; *American Journal of Sports Medicine*; *Journal of Bone and Joint Surgery*; *Arthroscopy*; *Bone & Joint Journal*; *Clinical Orthopaedics and Related Research*; *Knee Surgery, Sports Traumatology, Arthroscopy*; *Sports Health*; *Orthopaedic Journal of Sports Medicine*; and *Journal of Knee Surgery*. According to the 2018 InCites Journal Citation Reports index, these journals are recognized as the most impactful in the field of orthopaedic surgery, with impact factors of 11.645, 6.093, 4.716, 4.433, 4.301, 4.154, 3.149, 2.649, 2.589, and 1.591, respectively.

Inclusion and Exclusion Criteria

Three independent authors (C.B.E., A.J.C., E.S.C.) screened each search result to determine if it met inclusion and exclusion criteria. These were evaluated by the following inclusion criteria: (1) autografts with independent tunnel drilling techniques were implemented; (2) the patients underwent primary ACLR for chronic, subacute, or acute injuries; and (3) the study reported a 12-month minimum follow-up period. The studies were excluded if (1) the surgical technique was not explicitly stated, described, or referenced; (2) allografts or transtibial tunnel drilling techniques were implemented; (3) the patients underwent concomitant ligamentous repair or reconstructions at the time of ACLR, although partial meniscectomies and meniscus repairs were permitted; (4) the studies were on cadaveric, in vitro, or animal models; and (5) the studies utilized population databases, national registries, or cross-sectional data.

*Address correspondence to Cooper B. Ehlers, MD, MS, Georgetown University School of Medicine, 3900 Reservoir Rd NW, Washington, DC 20007, USA (email: cooper.ehlers@gmail.com).
†Georgetown University School of Medicine, El Segundo, California, USA.
‡MedStar Georgetown University Hospital, Washington, DC, USA.
§Inova Medical Group Orthopedics and Sports Medicine, Fairfax, Virginia, USA.
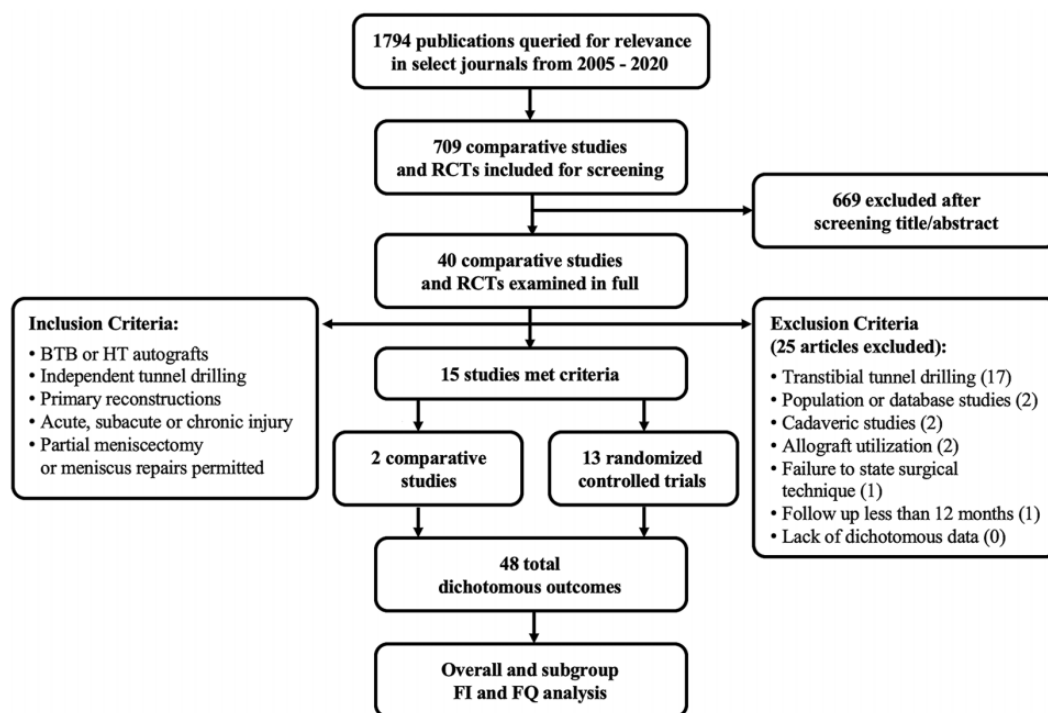
Figure 1. Study identification flowchart. BTB, bone–patellar tendon–bone; FI, fragility index; FQ, fragility quotient; HT, hamstring tendon; RCT, randomized controlled trial.

Risk-of-Bias Assessment and Methodology Scoring

Two authors (C.B.E., K.P.) independently evaluated each study. Risk of bias was assessed via the Cochrane Collaboration tool. Seven items were utilized to assess bias risk: random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), complete outcome data (attrition bias), selective reporting (reporting bias), and other bias. Scoring was determined using the signaling questions and algorithm provided by Cochran, with each category scored having a low, high, or unclear risk of bias. Methodology scoring was conducted according to the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) checklist.

Data Analysis

For each dichotomous outcome reported in a study, the following information was recorded: the type of outcome being measured, the number of patients in each outcome group, the population size, and the number of patients lost to follow-up. It was also recorded whether the outcome was listed as primary or secondary and if it was reported as significant or insignificant by recording the P value. Outcomes were considered primary if they were stated to be the primary outcome or if they were reported within the abstract, unless otherwise specified; all other outcomes were considered secondary. Reported P values were

verified for accuracy using the 2-tailed Fisher exact test. Statistical significance was defined as $P \leq .05$.

Through a trial-and-error method, outcome events were manipulated in a 2×2 contingency table until significance was reversed, as demonstrated in Figure 2. For example, if a particular outcome were initially reported as significant, the number of outcome event reversals required to raise the P value to $>.05$ was determined. Conversely, if the outcome was initially reported as nonsignificant, the number of outcome event reversals required to decrease the P value to $\leq .05$ was determined. The FI for each outcome was the number of event reversals necessary to alter statistical significance. The FQ was then determined for each outcome by dividing the FI by its respective sample size. For all included outcome events, mean FI and mean FQ were then determined, and interquartile ranges (IQRs) were calculated. A running total was kept of outcomes for which the FI was less than or equal to the number of patients lost to follow-up.

Three subgroups were analyzed for significant differences via independent t tests at 95% confidence: (1) primary versus secondary outcomes, (2) significant ($P \leq .05$) versus nonsignificant ($P > .05$) outcomes, and (3) outcomes for which the FI was less than the number of patients lost to follow-up versus outcomes for which the FI was greater than the number of patients lost to follow-up. Data analysis was performed in Microsoft Excel (Version 16.37).

The 48 dichotomous outcome measures ultimately included pivot-shift tests (n = 11), flexion/extension restrictions (n = 8), International Knee Documentation Committee (IKDC) ratings (n = 6), Lachman tests (n = 6), return to

Lachman Test	Positive	Negative	Patients Lost to Follow-Up
BTB	30	45	15
HT	*18*	*60*	12
P Value			0.04

Lachman Test	Positive	Negative	Patients Lost to Follow-Up
BTB	30	45	15
HT	*19*	*59*	12
P Value			0.056

Figure 2. Demonstration of fragility index = 1; a single-outcome event reversal resulting in altered statistical significance. BTB, bone–patellar tendon–bone; HT, hamstring tendon.

TABLE 1
Studies Meeting Study Inclusion Criteria (N = 15)[a]

Authors	Type of Study	Journal	Year	Mean FI	Mean FQ
Mayr et al[31]	RCT	*Arthroscopy*	2018	3.67	0.069
Mayr et al[32]	RCT	*Arthroscopy*	2016	4.00	0.143
Lao et al[26]	Comparative	*Arthroscopy*	2013	3.00	0.060
Fujita et al[13]	RCT	*Arthroscopy*	2011	2.33	0.064
Kim et al[24]	Comparative	*JBJS*	2009	2.25	0.037
Liu et al[28]	RCT	*AJSM*	2016	2.00	0.030
Karikis et al[21]	RCT	*AJSM*	2015	4.6	0.053
Ahldén et al[2]	RCT	*AJSM*	2013	4.25	0.043
Suomalainen et al[41]	RCT	*AJSM*	2012	3.00	0.068
Aglietti et al[1]	RCT	*AJSM*	2010	1.67	0.024
Gobbi et al[15]	RCT	*CORR*	2012	3.50	0.058
Karikis et al[20]	RCT	*KSSTA*	2017	4.67	0.050
Xu et al[47]	RCT	*KSSTA*	2014	2.50	0.038
Zaffagnini et al[48]	RCT	*KSSTA*	2011	2.50	0.032
Järvelä[19]	RCT	*KSSTA*	2007	2.76	0.049

[a]*AJSM, American Journal of Sports Medicine*; *CORR, Clinical Orthopaedics and Related Research*; FI, fragility index; FQ, fragility quotient; *JBJS, Journal of Bone and Joint Surgery*; *KSSTA, Knee Surgery, Sports Traumatology, Arthroscopy*; RCT, randomized controlled trial.

sport (n = 4), postoperative evidence of osteoarthritis (n = 4), rate of retear (n = 4), requirement for surgical reintervention (n = 2), the presence of anterior knee pain with kneeling (n = 2), and quality of bundle status at follow-up according to magnetic resonance imaging findings (n = 1). Nondichotomous data points were not included, as these cannot be analyzed with current fragility methodology.

RESULTS

Of the 1794 studies screened, 709 met initial search criteria, with 15 comparative studies ultimately included for analysis, 13 of which were RCTs. The included studies are detailed in Table 1. Nearly all studies utilized hamstring autografts, except for one study[19] that used single-bundle quadriceps tendon autografts and another study[47] that used single-bundle bone–patellar tendon–bone autografts.

A summary of the risk-of-bias assessment is shown in Figure 3, and methodology scoring is illustrated in Figure 4.

Overall and subgroup analyses of fragility are displayed in Table 2. Incorporating 48 total outcome events from all 15 studies, the overall mean FI was 3.14 (IQR, 2-4). The

overall FQ was 0.050 (IQR, 0.032-0.062). Of these 48 outcome events, 35 (72.9%) events had an associated FI that was less than the number lost to follow-up.

All 48 outcome events were recorded as either primary or secondary and as significant ($P \leq .05$) or nonsignificant ($P > .05$). In addition, each of the 48 outcomes reported an associated number of patients who were lost to follow-up. Accordingly, each outcome was classified according to these criteria, and 3 subgroupings were analyzed: primary versus secondary outcomes, significant versus insignificant outcomes, and outcomes for which the FI was less than the number of patients lost to follow-up versus outcomes for which the FI was greater than the number of patients lost to follow-up.

Significant (n = 7) and insignificant (n = 41) outcomes were analyzed and found to have medians FIs of 3 and 3, respectively. The mean FIs were 3.29 (IQR, 2-4) and 3.12 (IQR, 2-4), and the mean FQs were 0.047 (IQR, 0.029-0.069) and 0.051 (IQR, 0.033-0.060), respectively. There was no significant difference between mean FIs ($P = .77$; 95% CI, −0.974 to 1.301) or between mean FQs ($P = .60$; 95% CI, −0.013 to 0.023).

Primary (n = 37) and secondary (n = 11) outcomes were analyzed and found to have median FIs of 3 and 2,
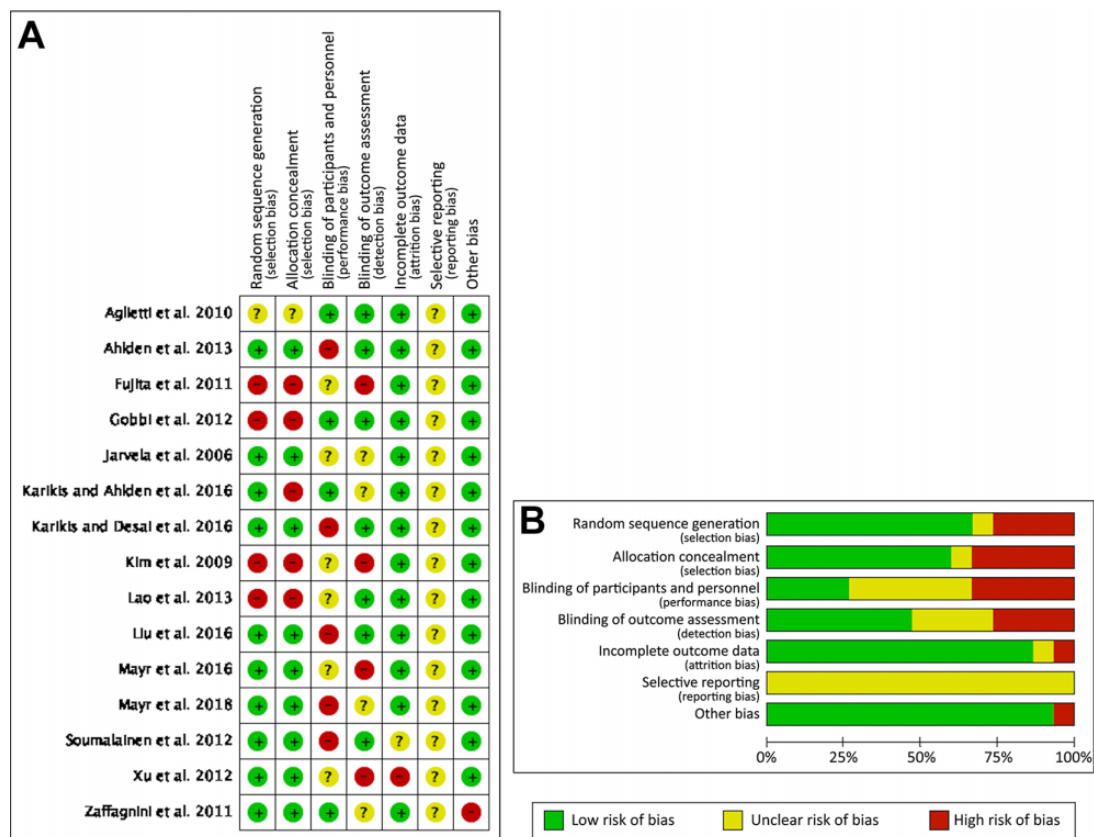
Figure 3. (A) Risk-of-bias assessment and (B) summary of risk-of-bias assessment according to the Cochrane Collaboration tool.
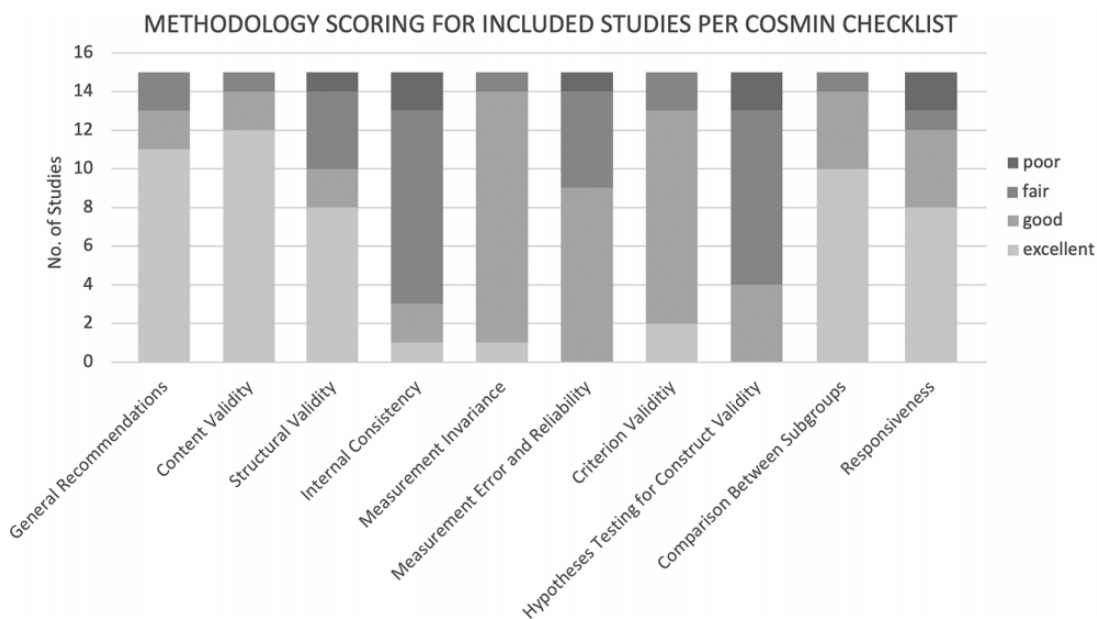

Figure 4. Summary of methodology scoring for the studies included by this systematic review. Following the layout of the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) checklist, the x-axis contains the categories within the checklist, and the y-axis depicts the number of included studies that fall in each category.

TABLE 2
Overall Fragility Data and Analysis of Subgroups[a]

Characteristic	Outcome Events	Fragility Index (IQR)	Fragility Quotient (IQR)
All trials	48	3.14 (2-4)	0.050 (0.032-0.062)
Outcome significance			
$P \le .05$	7	3.29 (2-4)	0.047 (0.029-0.069)
$P > .05$	41	3.12 (2-4)	0.051 (0.033-0.060)
P value		.77	.60
Outcome type			
Primary	37	3.32 (2-4)	0.053 (0.038-0.065)
Secondary	11	2.54 (2-3)	0.043 (0.028-0.052)
P value		.098	.178
Comparing outcome FI to LTF			
FI $\le$ LTF	35	3.40 (2-4)	0.054 (0.041-0.067)
FI $>$ LTF	13	2.46 (2-3)	0.041 (0.029-0.060)
P value		.033	.062

[a]$P \le .05$ represents the significant outcome subgroup, and $P > .05$ represents the nonsignificant outcome subgroup. FI $\le$ LTF represents the outcome subgroup for which the fragility index (FI) was less than the number of patients lost to follow-up (LTF), and FI $>$ LTF represents the outcome subgroup for which the FI was greater than the number of patients LTF.

respectively. The mean FIs were 3.32 (IQR, 2-4) and 2.54 (IQR, 2-3) and the mean FQs were 0.053 (IQR, 0.038-0.065) and 0.043 (IQR, 0.028-0.052), respectively. There was no significant difference between mean FIs ($P = .098$; 95% CI, –0.149 to 1.707) or between mean FQs ($P = .178$; 95% CI, –0.005 to 0.025).

For the outcomes where FI $\le$ LTF (n = 35), the median FI was found to be 3 and the mean FI to be 3.40 (IQR, 2-4). For the outcomes where FI $>$ LTF (n = 13), the median FI was found to be 2 and the mean FI to be 2.46 (IQR, 2-3). The associated mean FQs were 0.054 (IQR, 0.041-0.067) and 0.041 (IQR, 0.029-0.060), respectively. There was a significant difference between mean FIs ($P = .033$; 95% CI, 0.078-1.799) and no significant difference between mean FQs ($P = .062$; 95% CI, –0.001 to 0.027).

DISCUSSION

For this systematic review, the overall FI was found to be 3.14 and the overall FQ to be 0.050, which are findings consistent with prior orthopaedic literature reporting an average median FI of 2.5[11,17,22,23,30,44] and a mean FQ of 0.031.[35,44] Our findings demonstrate that statistical significance may be altered by the reversal of fewer than 4 outcome events or the reversal of 4% of outcome events. As hypothesized, the FI was less than the number of patients lost to follow-up in nearly three-quarters of outcomes (72.9%). We believe this is an important finding, as the composite results from comparative studies and RCTs are typically viewed as the best evidence available for influencing clinical practice and medical decision-making. Our

results emphasize the need to renovate classical statistical reporting.

P values with arbitrarily chosen α thresholds for significance are commonly utilized in medical and orthopaedic literature, despite well-known faults and shortcomings. *P* values are influenced by factors including effect size, sample size, and data dispersion.[40] Multiple studies[4,5,22] have shown that more than 50% of RCTs in sports medicine and orthopaedic literature do not report potential risk of bias, such as blinding of surgeons, patients, or outcome assessors. Further, up to 96% of abstracts and full-text biomedical articles report a minimum of 1 "statistically significant" result.[8] The inclusion of FI calculations can provide additional information, but there is no specific cutoff FI value that communicates the robustness of a study finding.[44] Generally, lower FI scores imply a weaker statistical strength, while higher FI scores imply statistical stability. However, reporting FI without corresponding FQ measurements has its pitfalls as well. Similar to the issue of isolated *P* value reporting, prior studies[17,23,30,44] indicate that FI does not inform on population size. The inclusion of FQ with FI and *P* values may help address this gap. Checketts et al[9] recently reviewed the robustness of clinical trials that were cited as having strong clinical evidence in the AAOS Clinical Practice Guidelines (CPG). The authors found the median FI and FQ were 2 and 0.022, respectively. The authors advocate for "triple reporting" of FI and FQ along with *P* values in the CPG recommendations to help clinicians understand the robustness of the individual trials that support specific recommendations.

Although this study directly examines fragility in the setting of ACLR, 1 prior study[42] did so indirectly through analysis of the Scandinavian knee ligament registries. These authors examined the fragility of 13 studies with median sample sizes of 5540, including large analyses of national databases.[14,36,37] The authors found the mean FI to be 178.5, with extensive variability (median, 116; range, 1-1089).[42] One possible explanation for the deviation of these values from prior fragility studies and from the results of our study is that Svantesson et al[42] reported an FI of zero in nearly one-third (30.4%) of the outcomes. An FI of zero indicates that zero outcome reversals were necessary to make a result insignificant because it was reported as a statistically insignificant finding; this is considered to be a "1-directional" fragility analysis. The authors exclusively examined the number of events necessary to make a significant result insignificant. In our analysis, we performed "2-directional" fragility analysis, as reported by Parisien et al.[35] This allows us to examine not only the number of event reversals required to make significant outcomes insignificant but also the number of event reversals needed to convert insignificant findings to significant. More outcomes are able to be examined through this technique, and it may allow for greater generalizability of findings.

Double-bundle ACLR allows for restoration of the 2 functional bundles of the ACL, the anteromedial and posterolateral bundles. The anteromedial bundle controls anteroposterior stability, while the posterolateral bundle primarily controls rotational stability. The principle behind this technique is to re-create the native ACL anatomy and

restore the proper tension pattern of each bundle. Although biomechanical studies[29,34] have shown the technique to be superior, many studies[7,10,25,27,29,43,46,49] have shown no significant difference with respect to subjective clinical outcomes. Despite the potential benefits of double-bundle ACLR, single-bundle anatomic ACLR remains the preferred technique for surgeons in the United States and globally.[16] This may be due to the technical demands of surgeries utilizing double-bundle grafts,[10] and the potential for increased difficulty of revision ACLR in these patients.[29] The inclusion of fragility analysis in future investigations may provide better clarity for graft bundle choice in ACLR.

This systematic review has several strengths. This study is strengthened by the utilization of 2-directional fragility analysis as discussed earlier. The study also examined primary and secondary outcomes to make our findings more generalizable. In addition to common primary outcomes such as retear rates and physical examination tests, many studies report radiologic findings or physical examination tests as secondary outcomes; our methodology captures all of these outcomes, allowing fragility to be applied more broadly. A final strength is the methodology of the literature search according to PRISMA guidelines. This search included orthopaedic journals with a mean impact factor of 5.04, which is higher than those of recent similar systematic reviews conducted on sports medicine $(3.2)^{22}$ and spine literature (2.4).[11]

However, this study is not without limitations. One potential limitation is the number of studies that met inclusion criteria is small in comparison with prior fragility analysis of medical and orthopaedic literature.[17,23,30,44] However, the scope of this systematic review was narrower, and therefore, it was more difficult to have a large number of articles meet our study inclusion. Another limitation is that FI and FQ can only evaluate categorical inputs with dichotomous outputs and therefore cannot be applied to continuous or ordinal variables. For example, rates of retear and physical examination maneuvers such as the Lachman or pivot-shift test results are easily encompassed, but pain measurements or functional outcomes scores cannot be captured. Other outcomes that were captured included return to play within a year and the presence or absence of radiographic findings. Additionally, some dichotomous outcome measures are inherently flawed. For example, return to sport may not be the strongest outcome measure as it is influenced by psychosocial factors as well as repair integrity. Third, this study did not track methods of graft fixation or other methods of bone tunneling. However, we believe the strict inclusion and exclusion criteria of our study are consistent with the current trends in surgical practices in the United States. Lastly, it is unknown if different levels of fragility exist between the high-impact journals included in our analysis and lower-impact journals and open access literature that were not included.

CONCLUSION

Studies comparing single-bundle versus double-bundle ACLR may not be as statistically stable as previously thought and may warrant additional investigation. Comparative studies and RCTs are at substantial risk for statistical fragility with few event reversals required to alter significance. The reversal of fewer than 4 outcome events in a treatment group can alter the statistical significance of a given result; this is commonly less than the number of patients lost to follow-up. Future comparative study analyses might consider including FI and FQ with P values in their statistical analysis.

REFERENCES

1. Aglietti P, Giron F, Losco M, Cuomo P, Ciardullo A, Mondanelli N. Comparison between single- and double-bundle anterior cruciate ligament reconstruction: a prospective, randomized, single-blinded clinical trial. *Am J Sports Med*. 2010;38(1):25-34. doi:10.1177/0363546509347096

2. Ahldén M, Sernert N, Karlsson J, Kartus J. A prospective randomized study comparing double- and single-bundle techniques for anterior cruciate ligament reconstruction. *Am J Sports Med*. 2013;41(11):2484-2491. doi:10.1177/0363546513497926

3. Ahmed W, Fowler RA, McCredie VA. Does sample size matter when interpreting the fragility index? *Crit Care Med*. 2016;44(11):e1142-e1143. doi:10.1097/CCM.0000000000001976

4. Bhandari M, Guyatt GH, Lochner H, Sprague S, Tornetta P. Application of the Consolidated Standards of Reporting Trials (CONSORT) in the fracture care literature. *J Bone Joint Surg Am*. 2002;84(3):485-489. doi:10.2106/00004623-200203000-00023

5. Bhandari M, Montori VM, Schemitsch EH. The undue influence of significant p-values on the perceived importance of study results. *Acta Orthop*. 2005;76(3):291-295. doi:10.1080/00016470510030724

6. Biau DJ, Jolles BM, Porcher R. P value and the theory of hypothesis testing: an explanation for new researchers. *Clin Orthop Relat Res*. 2010;468(3):885-892. doi:10.1007/s11999-009-1164-4

7. Björnsson H, Desai N, Musahl V, et al. Is double-bundle anterior cruciate ligament reconstruction superior to single-bundle? A comprehensive systematic review. *Knee Surg Sports Traumatol Arthrosc*. 2015;23(3):696-739. doi:10.1007/s00167-013-2666-x

8. Chavalarias D, Wallach JD, Li AHT, Ioannidis JPA. Evolution of reporting P values in the biomedical literature, 1990-2015. *JAMA*. 2016;315(11):1141-1148. doi:10.1001/jama.2016.1952

9. Checketts JX, Scott JT, Meyer C, Horn J, Jones J, Vassar M. The robustness of trials that guide evidence-based orthopaedic surgery. *J Bone Joint Surg Am*. 2018;100(12):e85. doi:10.2106/JBJS.17.01039

10. Desai N, Björnsson H, Musahl V, et al. Anatomic single- versus double-bundle ACL reconstruction: a meta-analysis. *Knee Surg Sport Traumatol Arthrosc*. 2014;22(5):1009-1023. doi:10.1007/s00167-013-2811-6

11. Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine J*. 2015;15(10):2188-2197. doi:10.1016/j.spinee.2015.06.004

12. Feinstein AR. The unit fragility index: an additional appraisal of "statistical significance" for a contrast of two proportions. *J Clin Epidemiol*. 1990;43(2):201-209. doi:10.1016/0895-4356(90)90186-S

13. Fujita N, Kuroda R, Matsumoto T, et al. Comparison of the clinical outcome of double-bundle, anteromedial single-bundle, and posterolateral single-bundle anterior cruciate ligament reconstruction using hamstring tendon graft with minimum 2-year follow-up. *Arthroscopy*. 2011;27(7):906-913. doi:10.1016/j.arthro.2011.02.015

14. Gifstad T, Foss OA, Engebretsen L, et al. Lower risk of revision with patellar tendon autografts compared with hamstring autografts: a registry study based on 45,998 primary ACL reconstructions in Scandinavia. *Am J Sports Med*. 2014;42(10):2319-2328. doi:10.1177/0363546514548164

15. Gobbi A, Mahajan V, Karnatzikos G, Nakamura N. Single- versus double-bundle ACL reconstruction: is there any difference in stability and function at 3-year followup? *Clin Orthop Relat Res*. 2012;470(3): 824-834. doi:10.1007/s11999-011-1940-9

16. Grassi A, Carulli C, Innocenti M, et al. New trends in anterior cruciate ligament reconstruction: a systematic review of national surveys of the last 5 years. *Joints*. 2018;6(3):177-187. doi:10.1055/s-0038-1672157

17. Grolleau F, Collins GS, Smarandache A, et al. The fragility and reliability of conclusions of anesthesia and critical care randomized trials with statistically significant findings: a systematic review. *Crit Care Med*. 2019;47(3):456-462. doi:10.1097/CCM.00000000 00003527

18. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294(2):218. doi:10.1001/jama. 294.2.218

19. Järvelä T. Double-bundle versus single-bundle anterior cruciate ligament reconstruction: a prospective, randomize clinical study. *Knee Surg Sport Traumatol Arthrosc*. 2007;15(5):500-507. doi:10.1007/s00167-006-0254-z

20. Karikis I, Ahldén M, Casut A, Sernert N, Kartus J. Comparison of outcome after anatomic double-bundle and antero-medial portal non-anatomic single-bundle reconstruction in ACL-injured patients. *Knee Surg Sport Traumatol Arthrosc*. 2017;25(4):1307-1315. doi:10.1007/s00167-016-4132-z

21. Karikis I, Desai N, Sernert N, Rostgard-Christensen L, Kartus J. Comparison of anatomic double- and single-bundle techniques for anterior cruciate ligament reconstruction using hamstring tendon autografts: a prospective randomized study with 5-year clinical and radiographic follow-up. *Am J Sports Med*. 2015;44(5): 1225-1236. doi:10.1177/0363546515626543

22. Khan M, Evaniew N, Gichuru M, et al. The fragility of statistically significant findings from randomized trials in sports surgery: a systematic survey. *Am J Sports Med*. 2017;45(9):2164-2170. doi:10.1177/0363546516674469

23. Khormaee S, Choe J, Ruzbarsky JJ, et al. The fragility of statistically significant results in pediatric orthopaedic randomized controlled trials as quantified by the fragility index: a systematic review. *J Pediatr Orthop*. 2018;38(8):e418-e423. doi:10.1097/BPO.0000 000000001201

24. Kim SJ, Chang JH, Kim TW, Jo SB, Oh KS. Anterior cruciate ligament reconstruction with use of a single or double-bundle technique in patients with generalized ligamentous laxity. *J Bone Jt Surg Ser A*. 2009;91(2):257-262. doi:10.2106/JBJS.H.00009

25. Kongtharvonskul J, Attia J, Thamakaison S, Kijkunasathian C, Woratanarat P, Thakkinstian A. Clinical outcomes of double- vs single-bundle anterior cruciate ligament reconstruction: a systematic review of randomized control trials. *Scand J Med Sci Sports*. 2013;23(1): 1-14. doi:10.1111/j.1600-0838.2011.01439.x

26. Lao ML, Chen JH, Wang CJ, Siu KK. Functional outcomes of y-graft double-bundle and single-bundle anterior cruciate ligament reconstruction of the knee. *Arthroscopy*. 2013;29(9):1525-1532. doi:10. 1016/j.arthro.2013.06.005

27. Li YL, Ning GZ, Wu Q, et al. Single-bundle or double-bundle for anterior cruciate ligament reconstruction: a meta-analysis. *Knee*. 2014. doi:10.1016/j.knee.2012.12.004

28. Liu Y, Cui G, Yan H, Yang Y, Ao Y. Comparison between single- and double-bundle anterior cruciate ligament reconstruction with 6- to 8-stranded hamstring autograft: a prospective, randomized clinical trial. *Am J Sports Med*. 2016;44(9):2314-2322. doi:10.1177/03635465 16650876

29. Mascarenhas R, Cvetanovich GL, Sayegh ET, et al. Does double-bundle anterior cruciate ligament reconstruction improve postoperative knee stability compared with single-bundle techniques? A systematic review of overlapping meta-analyses. *Arthroscopy*. 2015;31(6):1185-1196. doi:10.1016/j.arthro.2014.11.014

30. Matics TJ, Khan N, Jani P, Kane JM. The fragility of statistically significant findings in pediatric critical care randomized controlled trials. *Pediatr Crit Care Med*. 2019;20(6):e258-e262. doi:10.1097/PCC. 0000000000001922

31. Mayr HO, Benecke P, Hoell A, et al. Single-bundle versus double-bundle anterior cruciate ligament reconstruction: a comparative 2-year follow-up. *Arthroscopy*. 2016;32(1):34-42. doi:10.1016/j. arthro.2015.06.029

32. Mayr HO, Bruder S, Hube R, Bernstein A, Suedkamp NP, Stoehr A. Single-bundle versus double-bundle anterior cruciate ligament reconstruction—5-year results. *Arthroscopy*. 2018;34(9):2647-2653. doi:10.1016/j.arthro.2018.03.034.

33. Mohtadi NG, Chan DS. Return to sport-specific performance after primary anterior cruciate ligament reconstruction: a systematic review. *Am J Sports Med*. 2018;46(13):3307-3316. doi:10.1177/ 0363546517732541

34. Oh JY, Kim KT, Park YJ, et al. Biomechanical comparison of single-bundle versus double-bundle anterior cruciate ligament reconstruction: a meta-analysis. *Knee Surg Relat Res*. 2020;32(1):14-14. doi:10.1186/s43019-020-00033-8

35. Parisien RL, Dashe J, Cronin PK, Bhandari M, Tornetta P. Statistical significance in trauma research: too unstable to trust? *J Orthop Trauma*. 2019;33(12):e466-e470. doi:10.1097/BOT.0000000000 001595

36. Persson A, Fjeldsgaard K, Gjertsen JE, et al. Increased risk of revision with hamstring tendon grafts compared with patellar tendon grafts after anterior cruciate ligament reconstruction: a study of 12,643 patients from the Norwegian Cruciate Ligament Registry, 2004-2012. *Am J Sports Med*. 2014;42(2):285-291. doi:10.1177/03635 46513511419

37. Rahr-Wagner L, Thillemann TM, Pedersen AB, Lind M. Comparison of hamstring tendon and patellar tendon grafts in anterior cruciate ligament reconstruction in a nationwide population-based cohort study: results from the Danish registry of knee ligament reconstruction. *Am J Sports Med*. 2014;42(2):278-284. doi:10.1177/03635 46513509220

38. Renström PA. Eight clinical conundrums relating to anterior cruciate ligament (ACL) injury in sport: recent evidence and a personal reflection. *Br J Sports Med*. 2013;47(6):367-372. doi:10.1136/bjsports-2012-091623

39. Riboh JC, Hasselblad V, Godin JA, Mather RC. Transtibial versus independent drilling techniques for anterior cruciate ligament reconstruction: a systematic review, meta-analysis, and meta-regression. *Am J Sports Med*. 2013;41(11):2693-2702. doi:10.1177/03635465 13506979

40. Shochet LR, Kerr PG, Polkinghorne KR. The fragility of significant results underscores the need of larger randomized controlled trials in nephrology. *Kidney Int*. 2017;92(6):1469-1475. doi:10.1016/j.kint. 2017.05.011

41. Suomalainen P, Moisala AS, Paakkala A, Kannus P, Järvelä T. Double-bundle versus single-bundle anterior cruciate ligament reconstruction: randomized clinical and magnetic resonance imaging study with 2-year follow-up. *Am J Sports Med*. 2011;39(8): 1615-1622. doi:10.1177/0363546511405024

42. Svantesson E, Hamrin Senorski E, Danielsson A, et al. Strength in numbers? The fragility index of studies from the Scandinavian knee ligament registries. *Knee Surg Sport Traumatol Arthrosc*. 2020;28(2): 339-352. doi:10.1007/s00167-019-05551-x

43. Tiamklang T, Sumanont S, Foocharoen T, Laopaiboon M. Double-bundle versus single-bundle reconstruction for anterior cruciate ligament rupture in adults. *Cochrane Database Syst Rev*. 2012. doi:10.1002/14651858.cd008413.pub2

44. Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg*. 2019;154(1): 74-79. doi:10.1001/jamasurg.2018.4318

45. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a

fragility index. *J Clin Epidemiol*. 2014;67(6):622-628. doi:10.1016/j.jclinepi.2013.10.019

46. Xu M, Gao S, Zeng C, et al. Outcomes of anterior cruciate ligament reconstruction using single-bundle versus double-bundle technique: meta-analysis of 19 randomized controlled trials. *Arthroscopy*. 2013;29(2):357-365. doi:10.1016/j.arthro.2012.08.024

47. Xu Y, Ao Y, Wang J, Cui G. Prospective randomized comparison of anatomic single- and double-bundle anterior cruciate ligament reconstruction. *Knee Surg Sport Traumatol Arthrosc*. 2014;22(2):308-316. doi:10.1007/s00167-013-2398-y

48. Zaffagnini S, Bruni D, Marcheggiani Muccioli GM, et al. Single-bundle patellar tendon versus non-anatomical double-bundle hamstrings ACL reconstruction: a prospective randomized study at 8-year minimum follow-up. *Knee Surg Sport Traumatol Arthrosc*. 2011;19(3):390-397. doi:10.1007/s00167-010-1225-y

49. Zhu Y, Tang R, Zhao P, Zhu S, Li Y, Li J. Double-bundle reconstruction results in superior clinical outcome than single-bundle reconstruction. *Knee Surg Sport Traumatol Arthrosc*. 2013;21(5):1085-1096. doi:10.1007/s00167-012-2073-8