

The Statistical Fragility Of Clinical Trials Comparing Bone-Patellar Tendon-Bone and Hamstring Tendon Autografts In **Anterior Cruciate Ligament Reconstruction Surgery:** A Systematic Revie Jeffrey Okewunmi BS^a, Zakaria Chakrani BA^a John K. Cordero MD^a, Robert L. Parisien MD^a ^aIcahn School of Medicine at Mount Sinai, 1 Gustave L. Levy PI, New York, NY 10029





Disclosures:

Jeffrey Okewunmi, BS Nothing to Disclose

Zakaria Chakrani, BA Nothing to Disclose

John K. Cordero, MD Nothing to Disclose

Robert L. Parisien, MD

American Orthopaedic Society for Sports Medicine: Board or committee member Arthroscopy: Editorial or governing board Arthroscopy Association of North America: Board or committee member Journal of Cartilage & Joint Preservation: Editorial or governing board Journal of Sport Rehabilitation: Editorial or governing board New England Orthopaedic Society: Board or committee member Society of Military Orthopaedic Surgeons: Board or committee member



Background

- Physicians rely on P-values when interpreting clinical trial data. However, this metric neglects loss to follow-up, sample size, and power.¹⁻³
- Statistical fragility assesses the robustness of clinical data based on the number of outcome event reversals required to reverse statistical findings, and may address the limitations of the P-value.⁴
 - **Fragility index (FI)** represents the number of outcome reversals required to switch a statistically significant result into a non-significant result
 - **Reverse fragility index (reverse FI)** represents the number of outcome reversals required to switch a statistically significant result into a non-significant result
 - **Fragility Quotient (FQ)** accounts for sample size by dividing FI or reverse FI by sample size

Hamstring tendon (HT) and the bone-patellar tendon-bone (BPTB) are the two most common autograft choices for ACL reconstruction. However there is no current consensus on the which is clinically superior.⁵



Study Aim & Hypothesis

This study aims to assess the statistical fragility of recently published, clinical studies that compare Bone-Patellar Tendon-Bone (BPTB) and Hamstring Tendon (HT) autografts In ACL reconstruction surgery

We hypothesized that a fragility analysis of RCTs investigating autograft choice in ACL reconstruction surgery would demonstrate fragility, with few outcome event changes required to reverse statistical significance



Methods

- Retrospective review querying PubMed/MEDLINE/EMBASE for articles from 1/2010 to 2/2021
 - Inclusion criteria: clinical trials that reported on patients undergoing primary, unilateral ACL reconstruction surgery stratified by autograft type (HT vs BPTB). The studies reported at least one dichotomous, categorical outcome and were available in the English language.
 - Exclusion criteria: systematic reviews, non-clinical trials, greater than two treatment groups, revision surgery, and/or reported in vitro, cadaveric, or animal data.
- Extracted data included sample size, number of outcomes from each intervention group, loss to follow-up, and P-values
- FI and reverse FI were calculated using a 2 x 2 contingency table, by manipulating the outcome events until reversal of significance. FQ was calculated by dividing the FI or reverse FI by sample size

	+ Outcome	- Outcome
НТ	2	98
ВРТВ	10	90
		<i>P</i> -value 0.033

	+ Outcome	- Outcome
НТ	3	97
ВРТВ	10	90
		<i>P</i> -value 0.082

Results

- Search yielded 208 total articles, with 127 undergoing full-text review
- 26 studies met inclusion criteria
- 102 total dichotomous outcomes:
 - 10 significant (P<0.05)
 - 92 non-significant (P>0.05)





Identification of studies via databases and registers

Records removed *before screening*: Duplicate records removed (n = 91)

Records excluded (n = 67)

Reports not retrieved (n = 0)

Reports excluded (n = 34) More than 2 treatment groups In vitro, animal, cadaveric data Revision surgery Non-dichotomous outcome No English translation

Results

- 102 total outcomes: median FI 5 (IQR 3 to 6); median FQ 0.057 (IQR 0.033 to 0.096)
- 10 significant outcomes: median FI 3 (IQR 2 to 5); median FQ 0.018 (IQR 0.017 to 0.039)
- 92 non-significant outcomes: median reverse FI 5 (IQR 4 to 6); median FQ 0.064 (IQR 0.038 to 0.101)
- Substantial statistical fragility observed across outcome categories

	Events	FI (IQR)	FQ (IQR)
All Outcomes	102	5 (3 to 6)	0.057 (0.033 to 0.096)
Reported P-Value			
<0.05 (statistically significant)	10	3 (2 to 5)	0.018 (0.017 to 0.039)
>0.05 (not statistically significant)	92	5 (4 to 6)	0.064 (0.038 to 0.101)
Outcome Category			
Lachman Test	13	6 (5 to 8)	0.062 (0.041 to 0.122)
Pivot Shift Test	10	5 (4 to 7)	0.054 (0.040 to 0.068)
Extension Deficit	7	5 (4.5 to 8)	0.077 (0.034 to 0.083)
Flexion Deficit	7	6 (4.5 to 6.5)	0.041 (0.039 to 0.049)
Knee Pain	11	5 (3 to 5.5)	0.058 (0.021 to 0.113)

3 to 0.096) 0.017 to 0.039) Q 0.064 (IQR 0.038 to

Limitations

- Given narrow topic of investigation, our analysis included a small number of eligible studies and a limited amount of statistically significant outcomes for evaluation
- Fragility analysis is limited to dichotomous, categorical outcomes, and is not generalizable to continuous variables
- Standardized FI/FQ thresholds for evaluating trial data have not been established. It is unclear how these results should impact clinical decision making



Conclusions

- RCTs evaluating autograft choice in ACL reconstruction demonstrate substantial fragility, as the • reversal of a small number of outcomes is sufficient to alter statistical significance
- Statistical fragility is higher when considering outcomes reported as statistically significant
- We therefore recommend the inclusion of a comprehensive fragility analysis (FI and FQ), in addition to the P-value, to allow for increased reliability in the interpretation of literature pertaining to ACL reconstruction



References

- Sterne JAC, Smith GD. Sifting the evidence-what's wrong with significance tests? *Phys Ther*. 1. 2001;81(8):1464-1469.
- Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. JAMA. 2. 2005;294(2):218-228.
- Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled 3. trial results is frequently fragile: a case for a Fragility Index. J Clin Epidemiol. 2014;67(6):622-628.
- Feinstein AR. The unit fragility index: an additional appraisal of "statistical significance" for a 4. contrast of two proportions. J Clin Epidemiol. 1990;43(2):201-209.
- 5. Samuelsen BT, Webster KE, Johnson NR, Hewett TE, Krych AJ. Hamstring Autograft versus Patellar Tendon Autograft for ACL Reconstruction: Is There a Difference in Graft Failure Rate? A Meta-analysis of 47,613 Patients. Clin Orthop Relat Res. 2017;475(10):2459-2468.

