Internet Medical Journal
ISSN 1093-7935 · medjournal.com · Est. 1996

Reverse Continuous Fragility Index Misleads in Biceps Tenotomy Versus Tenodesis Trials: The Case for a Continuous Fragility Score and Quotient

Author(s)	Thomas F. Heston
Affiliation(s)	Department of Family Medicine, University of Washington, Seattle, USA
Affiliation(s)	Department of Medical Education and Clinical Sciences, Elson S. Floyd College of Medicine, Washington State University, Spokane, USA
ORCID	0000-0002-5655-2512

Published	23 APR 2026
DOI	https://doi.org/10.5281/zenodo.19720537
Article type	Commentary
Citation	Heston TF. Reverse Continuous Fragility Index Misleads in Biceps Tenotomy Versus Tenodesis Trials: The Case for a Continuous Fragility Score and Quotient. Internet Medical Journal. 2026;1:e19720537

Abstract

Shoulder surgeons regularly choose between biceps tenotomy and tenodesis during rotator cuff repair, and that choice is guided by evidence synthesis built on fragility analysis of the underlying randomized controlled trials. A recent analysis of seven such trials reports a mean reverse continuous fragility index (rCFI) of 17.7 and concludes that the noninferiority findings are moderately stable — a conclusion that flows directly into how surgeons counsel patients and how guidelines grade recommendations. The rCFI calculation simulates a reconstructed patient-level data set from published summary statistics and returns an iteration count that varies with the random seed; its normalization, labeled the "reverse fragility quotient," divides simulated transfers by the number of real patients. A measurement that shifts with investigator choices and mixes simulated with real quantities cannot serve as a foundation for clinical decisions. The continuous fragility quotient (CFQ) is a structurally different construction: it computes a deterministic value on the interval from zero to one directly from the Welch t-statistic using the same published inputs, with no simulation involved. Replacing rCFI with CFQ gives surgeons, patients, and guideline

committees a reproducible measurement they can rely on — a foundational requirement for honest evidence synthesis and sound clinical recommendations.

Keywords

statistical fragility, continuous fragility quotient, reverse continuous fragility index, noninferiority trials, biceps tenodesis, shoulder arthroscopy, evidence-based medicine

Biceps tenotomy and tenodesis are common procedures performed during shoulder arthroscopy, and a recent systematic review of seven randomized controlled trials concludes that the evidence supports the noninferiority of one procedure compared with the other, with moderate stability, anchored by a mean reverse continuous fragility index (rCFI) of 17.7 (1). Because that conclusion will inform shared decision-making with patients and the development of future clinical guidelines, the method underlying the rCFI value warrants scrutiny. Two features of the rCFI do not meet the standards that clinical evidence synthesis requires: the calculation produces different values on repeat computation of the same trial, and its normalized form — labeled "reverse fragility quotient" in the source analysis — does not share the mathematical properties of the fragility quotients established for binary outcomes. A continuous-outcome fragility measure that preserves those properties already exists: the continuous fragility quotient (CFQ), defined algebraically from the Welch t-statistic geometry. Replacing rCFI with CFQ yields a reproducible measurement that surgeons and guideline committees can rely on.

Fragility analysis exists to answer a specific clinical question: how much would a trial's data need to change before the statistical conclusion flips? A fragility quotient gives a proportion-based answer — what fraction of the sample would need to change — and for that proportion to mean anything across trials, across specialties, and across journals, it must be computed deterministically from the observed evidence without simulation, reconstruction, or distributional assumption. The original fragility index for binary outcomes laid the foundation by counting real outcome changes in real patients (2), and this count was subsequently normalized by the number of real patients to create the proportion-based Fragility Quotient (3). Subsequent extensions — the modified-arm fragility quotient (MFQ) for allocation-imbalanced trials and the global fragility quotient (GFQ) for multi-category outcomes — preserved the essential property that the numerator and denominator describe the same real cohort. A continuous-outcome fragility quotient ought to preserve that property on its own terms: different mathematical machinery, same interpretive discipline (4). Furthermore, the source analysis labels its primary conclusion 'moderate robustness,' but the rCFI measures how far the data would need to shift before the significance classification flips — which is a fragility property (classification stability), not a robustness property. True robustness in the statistical-evidence sense measures geometric

distance from therapeutic neutrality, a separate dimension that the rCFI does not register (5).

The recent biceps-trials analysis departs from those standards in two specific ways, each with direct bearing on whether the reported rCFI values can support the moderate-stability conclusion. First, rCFI relies on a stochastic simulation method that generates a pseudo-individual data set from the published means, standard deviations, and sample sizes under distributional assumptions, and then iteratively transfers data points between the two simulated groups until a Welch t-test crosses $\alpha = 0.05$. The underlying continuous fragility index (CFI) algorithm was originally defined to iterate a significant result toward nonsignificance (6); rCFI is the reverse application, iterating a nonsignificant result toward significance, with the iteration count reported as the index value.

Because the simulated data set depends on the random seed, rCFI is stochastic: running the same calculation again on the same trial with a different random seed produces a different value. The source analysis runs five simulations per trial "to address the risk of random error," which concedes the instability rather than eliminating it (1). Second, the normalization labeled "reverse fragility quotient" in the same analysis divides rCFI by the actual trial sample size. The binary-outcome FI/N convention that this label imports has an internally consistent unit structure — real patient toggles in the numerator, real patient count in the denominator. The continuous analog breaks that consistency: the numerator counts transfers of simulated datapoints while the denominator counts real trial participants. A measurement that shifts with investigator choices and mixes simulated with real quantities cannot carry the weight that a moderately stable conclusion places on it.

An alternative to the CFI that does not rely on stochastic simulations and distribution assumptions is the Continuous Fragility Score (CFS) and its quotient, the CFQ. The CFS takes the same inputs — the mean of group 1, the mean of group 2, their standard deviations, and sample sizes — and computes fragility algebraically from Welch t-statistic geometry. The observed Welch t-statistic T is calculated directly; the critical value t^* at $\alpha = 0.05$ is determined from the Welch-Satterthwaite degrees of freedom; the continuous fragility score (CFS) = $| |T| - t^* |$ is the standard-error-unit distance from the $\alpha = 0.05$ boundary; and CFQ = CFS / (1 + CFS) maps that distance onto the zero-to-one proportion scale for normalization and cross-trial systematic reviews and meta-analyses.

The CFS has clear advantages over the CFI: a) no stochastic simulations are performed as the CFS is calculated directly from the outcome data; b) no distributional reconstruction is invoked; c) the value is deterministic: any investigator running the same calculation on the same published summary statistics obtains the same CFQ, every time; d) the same algebraic computation principle extends to multi-group, correlation, ordinal, and survival designs through parallel fragility quotients, placing continuous-outcome fragility on the same footing as the binary and multinomial cases.

A paired robustness metric, the Meaningful Change Index (MeCI), quantifies how far the nearer group mean sits from the distributional crossover point between the two groups, normalized by their combined standard deviations — providing a p-value-independent

measure of group distinguishability (7). Together, CFQ and MeCI supply reproducible measurements of both classification stability and distance from therapeutic neutrality — two core dimensions that, alongside the p-value, constitute complete statistical evidence for any claim of effect or non-effect.

The biceps trial set provides the summary statistics required for CFQ computation, since the same inputs drive the rCFI calculations reported in the source analysis. Two trials in the set illustrate why reproducible fragility and robustness metrics would yield a different clinical picture than the rCFI-based conclusion. The largest trial (n = 151) returned a reverse fragility quotient of 0.053 (8); the source analysis attributes this value to "smaller between-group differences" — which is precisely what proximity to therapeutic neutrality means clinically, and what a dedicated robustness metric such as the MeCI would display directly rather than bury in an averaged summary.

A second trial yielded a reverse fragility quotient of 0.400, a 7.5-fold increase over the first within the same metric (9). The arithmetic mean of these values, 0.227, hides a bimodal distribution that a surgeon or guideline committee working from the summary alone cannot see. Separately, the same analysis applies the classic fragility index to the biceps deformity outcome across four included trials, reports a mean fragility index of 5 (range 1 to 15), and explicitly characterizes those findings as relatively fragile. Hu's own secondary analysis, therefore, demonstrates fragility in a subset of the same cohort whose primary outcomes the rCFI characterizes as moderately stable — a contradiction that a unified system of correctly unit-matched metrics (using the continuous CFQ for the primary scores and the binary MFQ for the deformity outcome) would have flagged from the outset.

A reasonable response is to adopt reproducible metrics of fragility and robustness for continuous-outcome noninferiority analyses in orthopedic sports medicine. CFQ and MeCI require only the same published summary statistics that rCFI already uses, so the added reporting burden is computational rather than evidentiary. The resulting measurements are reproducible across investigators and aligned with the binary and multinomial fragility quotients already used in the sports-medicine literature for dichotomous outcomes. CFQ is computed directly from reported summary statistics without simulation or distributional reconstruction; MeCI uses the same summary statistics and assumes approximately normal group distributions for its distributional crossover-point calculation. Reporting these dimensions together would give a surgeon complete statistical evidence: a) the p-value, representing the compatibility of the observed data with the null hypothesis of no effect; b) the fragility of the p-value-derived significance classification, measuring how close the observed test statistic sits to the significance boundary in units of sampling uncertainty; and c) the robustness of the results, measuring how distinguishable the two patient populations are based on their distributional overlap, independent of the sample size used to observe them.

Giving clinicians the significance classification, its fragility, and the underlying population distinguishability provides complete statistical evidence. Until clinical trials report all three metrics, with fragility and robustness kept analytically separate, interpretive paradoxes of

the sort illustrated by the biceps trial set will continue to mislead; as a result, surgical decisions, patient conversations, and guideline recommendations will be determined by incomplete and often misleading statistical evidence.

Declarations

Funding: This study did not receive any external funding.

Conflicts of Interest: The author reports no conflicts of interest.

Data Availability: Not applicable.

Research Ethics Statement: Not applicable. This commentary did not involve human subjects research, animal research, or protected health information.

AI Usage: Large language models were used for language editing and formatting assistance; the author reviewed, verified, and is fully responsible for all content.

References

1. Hu EY, Althoff AD, Cervantes JE, Dave U, Kessler KG, Moran TE. Statistical Robustness of Randomized Controlled Trials Comparing Biceps Tenotomy Versus Tenodesis: A Reverse Continuous Fragility Index Analysis. Am J Sports Med. 2026 Apr 22;3635465261440392. doi:10.1177/03635465261440392 PubMed PMID: 42015691.

2. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. J Clin Epidemiol. 2014 Jun;67(6):622–8. doi:10.1016/j.jclinepi.2013.10.019

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

4. Heston TF. Significance, Fragility, and Robustness in Clinical Trials: Stratifying Statistical Evidence. Cureus. 2025 Dec 31;17(12). doi:10.7759/cureus.100494

5. Heston TF. The Neutrality Boundary Framework: Quantifying Statistical Robustness Geometrically. arXiv. 2025 Nov 2;2511.00982. doi:10.48550/arXiv.2511.00982

6. Caldwell JME, Youssefzadeh K, Limpisvasti O. A method for calculating the fragility index of continuous outcomes. J Clin Epidemiol. 2021 Aug;136:20-25. doi:10.1016/j.jclinepi.2021.02.023

7. Heston TF. Meaningful Change Index: A P-Value Independent Metric for Assessing Robustness and Fragility in Continuous Outcomes. Zenodo. 2025 Sep 26. doi:10.5281/zenodo.17212383

8. Zhang Q, Zhou J, Ge H, Cheng B. Tenotomy or tenodesis for long head biceps lesions in shoulders with reparable rotator cuff tears: a prospective randomised trial. Knee Surg Sports Traumatol Arthrosc. 2015 Feb 1;23(2):464–9. doi:10.1007/s00167-013-2587-8

9. Castricini R, Familiari F, De Gori M, Riccelli DA, De Benedetto M, Orlando N, et al. Tenodesis is not superior to tenotomy in the treatment of the long head of biceps tendon lesions. Knee Surg Sports Traumatol Arthrosc. 2018 Jan 1;26(1):169–75. doi:10.1007/s00167-017-4609-4