The Continuous Fragility Quotient as a Model-Free Assessment for Continuous Outcomes

Author: Thomas F. Heston, MD, MSc

Affiliations: Department of Family Medicine, University of Washington, Seattle, WA; Department of Medical Education and Clinical Sciences, Washington State University, Spokane, WA

ORCID: 0000-0002-5655-2512

Published: Internet Medical Journal 2026;1(1). doi:10.5281/zenodo.19445024

The continuous fragility index (CFI) extended fragility assessment to continuous outcomes, but involved simulating candidate datasets when original data are unavailable [1]. This reconstruction step imports model-dependent assumptions that make the fragility measure depend on modeling choices, turning it into a property of the chosen model rather than of the observed evidence. The continuous fragility quotient (CFQ), a model-free alternative, resolves this problem by deriving fragility entirely from the test geometry that yields the reported *p* value, yielding a uniquely reproducible value from summary statistics alone [2].

Statistical fragility quantifies the minimal perturbation required to reverse a significance classification, measured either as changes to observed outcomes in exact discrete designs or as movement in test-statistic space for continuous designs such as the CFQ. For binary outcomes, this is operationalized as a raw count — the fragility index (FI) for two-arm trials [3] or the global fragility index (GFI) for any r×c contingency table [4] — representing the minimum outcome changes needed to cross the $p = 0.05$ boundary using only the observed table and exact tests. These counts are then normalized to the sample size to produce fragility quotients, enabling cross-study comparison. The model-free property of these metrics is not incidental; it is what makes fragility a verifiable, analyst-independent property of the evidence.

The continuous fragility index (CFI) extended fragility assessment to continuous outcomes by using an iterative substitution algorithm to calculate fragility prospectively from data or retrospectively from summary statistics. The method thus imports a specific distributional assumption (normality) and a reconstruction algorithm whose details are not uniquely determined by the reported summaries. Two analysts given identical means, standard deviations, and sample sizes can produce different CFI values depending on the reconstruction seed, substitution rule, and convergence criteria applied. A recent commentary advocating the CFI as a complement to the binary fragility index left this model-dependency unaddressed [5]. The field is therefore working with a continuous fragility metric whose value is not uniquely determined by the evidence it purports to characterize — precisely the property that fragility assessment exists to avoid.

The CFQ resolves this by operating entirely within the geometry of the Welch *t* test when the reported *p* value arises from, or can be validly represented in, Welch *t*-statistic form for an independent two-group mean comparison. Given group means (m_1, m_2), standard deviations (s_1, s_2), and sample sizes (n_1, n_2), the observed Welch t-statistic T and the critical value t* from the appropriate t-distribution determine a unique SE-scaled distance to the significance boundary — the continuous fragility score, CFS = ||T| − t*|. The CFQ maps this distance to the unit interval: CFQ = CFS / (1 + CFS). No observations are reconstructed, no distributional family is assumed beyond the Welch structure already implicit in the reported *p* value, and every analyst computing CFQ from the same summary statistics produces an identical value. The distinction from the CFI is structural: the CFI is a model-dependent count that requires reconstruction; the CFQ is a model-free quotient that requires only the test statistic and its critical value [2].

Three properties confirm the advantages of CFQ. First, because CFQ is derived from the same test statistic that generates the *p* value, internal consistency between the significance decision and the fragility assessment is guaranteed by construction — no test-statistic mismatch can introduce spurious variation. Second, CFQ pairs directly with the Meaningful Change Index, MeCI = |T| / (1 + |T|), a robustness metric quantifying geometric distance from therapeutic neutrality using the same Welch t-statistic, completing the p–fr–nb triplet — significance, fragility, and robustness — from summary statistics without additional modeling [2]. Third, CFQ is normalized to [0, 1] and increases monotonically as the t-statistic moves away from the significance boundary in either direction, enabling

direct cross-study comparison. The CFI, as an unnormalized count, is sample-size dependent, complicating cross-study comparisons without additional normalization.

The appropriate metric for continuous outcome fragility is the CFQ: model-free, uniquely determined by reported summary statistics, normalized for cross-study comparison, and paired with a complementary robustness measure within a unified evidence framework. The CFI identified a genuine need — continuous outcomes had no fragility metric — but addressed it with a method that introduces exactly the analyst-dependent variation that fragility assessment exists to eliminate. Journals publishing fragility analyses of continuous-outcome trials should consider requiring model-free metrics to ensure that reported fragility values characterize evidence properties rather than method choices. Fragility that varies by analyst is not a measure of evidence quality — it is a measure of assumption quality.

Declaration

The author reports no conflicts of interest. This study did not receive any external funding. Large language models were used for language editing and formatting assistance; the author reviewed, verified, and is fully responsible for all content.

References

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

2. Heston TF. Fragility Metrics Toolkit. Zenodo; 2026. doi:10.5281/zenodo.17254763

3. 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;67: 622–628. doi:10.1016/j.jclinepi.2013.10.019

4. Heston TF. The Global Fragility Index: A Path-Independent Measure of Statistical Fragility. Zenodo. 2025 [cited 4 Apr 2026]. doi:10.5281/zenodo.18078509

5. Chan J, Vrla M, Thompson C, Trofa D, Li X, Wang D, et al. Fragility of Assumptions: Response. Orthop J Sports Med. 2026;14: 23259671251409149. doi:10.1177/23259671251409149