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

From Fragility Index to Fragility Distance in Clinical Trials

Thomas F. Heston

Affiliation	Department of Family Medicine, University of Washington, Seattle, USA
Affiliation	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 Aug 2026
DOI	10.5281/zenodo.22072163
Article type	Commentary
Citation	Heston TF. From Fragility Index to Fragility Distance in Clinical Trials. Internet Medical Journal. 2026;1:e22072163

Abstract

Fragility indexes attempt to quantify the minimum perturbation that reverses a sample's statistical significance under an unstated constraint: the trial roster is closed, so only the recorded outcomes of enrolled patients may change. The fragility distance (FD) removes that constraint, defining fragility as the L1 edit distance from the observed 2×2 table to the significance boundary, where one unit is one patient added to or removed from any cell. Because every constrained index bounds the FD from above, the published toggle and transfer indexes become special cases of a single geometric quantity. In a published phase 2 trial reporting no significant difference ($p = 0.0903$), FD = 1: the removal of a single patient reverses the significance classification, a proximity to the boundary that the global fragility index, also equal to 1 but denominated in two-cell moves, understates by half. Reporting FD with its witness table gives the true minimum perturbation, priced in the one unit that occurs in every real trial: the enrollment, or the loss, of a single patient.

Keywords

fragility distance, fragility index, global fragility index, statistical fragility, perturbation fragility, edit distance, clinical trials

Statistical fragility metrics for two-arm binary outcome trials, dating from the unit fragility index in 1990 to the present, attempt to quantify how unstable a significance classification is based on recategorizing patients within the sample while keeping the trial roster closed. Although fragility indexes and their extensions have the goal of measuring the minimum perturbation that reverses a significance classification, each measures that minimum only within a self-imposed constrained move set. The unit fragility index proposed toggling outcomes such that both row and column marginal totals remained fixed [1,2]. The Walsh Fragility Index (FI) toggles patients from a non-event to an event within a single arm, resulting in fixed row marginal totals, unfixed column marginal totals, and a fixed sample size [3]. The Global Fragility Index allows transfers from any cell to any other cell with no requirement to keep marginal totals fixed, although total sample size remains fixed [4]. However, the true minimum perturbation to flip the statistical significance of a contingency table requires no constraints on marginal totals or total sample size. When marginal totals and sample size are unconstrained, the minimum number of single-patient moves is the L1 edit distance from the observed 2×2 table to the nearest table on the other side of the significance boundary. This is the "taxicab" distance from the observed sample distribution to the closest distribution where the significance classification changes, i.e., where the P-value crosses the alpha threshold of 0.05. This unconstrained perturbation is the Fragility Distance (FD), in which the unit of perturbation to the contingency table consists of one patient added to, or removed from, any cell.

Fragility indexes quantify *perturbation* fragility, which asks how far the observed table sits from the decision boundary where the P-value crosses 0.05. Perturbation fragility is a geometric quantity that evaluates the sample itself. This is distinct from *resampling* fragility, which asks how often a new sample of the same size would flip statistical significance; because it derives from the same observed result, resampling fragility is strongly associated with the P-value but not identical to it [5]. While the fragility index was originally proposed to be the minimum perturbation of the contingency table to flip statistical significance (as determined by the P-value), the true minimum perturbation is the L1 edit distance itself; every constrained index can only equal or exceed it.

The space of two-arm binary results is the set of all 2×2 tables with nonnegative whole-number cells and any total enrollment, and the L1 edit distance defined above is the shortest distance between any two tables in that space. A toggle switches one patient's outcome, and a transfer moves one patient between any two cells; both toggles and transfers change two cells at once, and therefore each costs two L1 units. It follows that the toggle and transfer counts are path lengths measured along restricted routes, not distances in the underlying

space. The restriction, moreover, prices the wrong doubt: nothing in clinical reality holds enrollment fixed, because trials routinely enroll one more patient or lose one to follow-up. The nearest published construct reasons about the unobserved outcomes of patients lost to follow-up, but only for patients already enrolled, and as a sensitivity analysis rather than a minimum [6]. The category of a true distance to the significance boundary has remained unoccupied.

FD, the fragility distance, quantifies this shortest distance from the observed contingency table to the nearest table with the opposite significance classification. FD is the minimum number of unit edits, each adding one patient to any cell or removing one patient from any cell, that moves the observed table across p = 0.05 under the two-sided Fisher's exact test; total enrollment, row totals, and column totals are all free, and the definition applies to significant and nonsignificant baselines alike. Because every toggle and every transfer is itself a pair of edits, the toggle and transfer indexes bound the distance from above: FD ≤ 2·GFI ≤ 2·FI always holds, since removing a restriction can never lengthen the shortest path, and adding one can never shorten it. The coupled move of the unit fragility index changes four cells at once, so its bound is FD ≤ 4·UFI.

In some cases, the addition-only and deletion-only fragility distances (aFD and dFD) may be used: aFD = k means that k additional patients, with particular outcomes in particular arms, could have changed the trial's statistical conclusion, and dFD = k means the loss of k particular patients could have done the same. These constrained fragility distances bound FD under their respective mechanisms, and because they are constrained, FD ≤ aFD and FD ≤ dFD. The quotient form is the fragility distance quotient, FDQ = FD/N, where N is the observed enrollment. FDQ supports cross-trial comparison, although it is not numerically interchangeable with FQ or GFQ, because a reclassification costs two edits on the FDQ scale but one move, a toggle or a transfer, on the FQ/GFQ scale.

Published trials make the geometry concrete, and the arithmetic can be checked by hand from the four cell counts. In a phase 2 trial of semaglutide versus placebo in biopsy-proven non-alcoholic steatohepatitis (NASH) related cirrhosis, 5 of 47 patients in the semaglutide group showed improvement compared to 7 of 24 in the placebo group (reported p=0.087; two-sided Fisher exact test on the counts, p=0.0903). The FD = 1 by three distinct edits: a) removing a single semaglutide-group responder, for example to loss of follow-up, changes the Fisher exact P-value to 0.0381, and the significance classification reverses, with the difference favoring placebo now crossing the 0.05 threshold {5, 42, 7, 17 → 4, 42, 7, 17}; b) add one placebo responder {5, 42, 8, 17} (p = 0.0498); or c) remove one placebo non-responder {5, 42, 7, 16} (p = 0.0498). The authors concluded that, in this patient population, semaglutide did not significantly improve fibrosis or achievement of NASH resolution [7]. This headline conclusion, however, conceals that the classification relies on just one patient, and that the nearest significant table lies in the direction favoring placebo. A reader told only

"no significant difference" cannot see that the evidence sits one edit from the opposite conclusion ("a significant difference"). While the FI and GFI also equal 1, one toggle or transfer requires two table changes, one cell +1 and another cell -1, so an index of 1 corresponds to two single-patient edits. FD = 1 shows the table sits half that distance from the boundary: the fragility is greater than the FI and GFI suggest.

The practical step is modest: trialists and meta-researchers reporting fragility for a 2×2 result should report the FD together with its witness table, the edited table that attains the minimum, which shows the reader exactly which patients the verdict depends on. Future research on the distributions of FD and FDQ across published trial corpora will help identify interpretive bands; however, as a provisional convention pending that empirical work, an FDQ of 0.05 or less may be taken to suggest an unstable (fragile) significance classification.

FD measures how far the evidence actually sits from the boundary that changes a research trial's significance interpretation. For thirty-five years, fragility research has measured the distance to the significance boundary with a ruler forbidden to leave the trial; FD identifies a study's fragility more clearly by removing arbitrary constraints.

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

2. Walter SD. Statistical significance and fragility criteria for assessing a difference of two proportions. J Clin Epidemiol. 1991;44: 1373–1378. doi:10.1016/0895-4356(91)90098-T

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. Bidirectional fragility is a step forward but not far enough: the case for a global fragility index. Internet Med J. 2026;1: e19464166–e19464166. doi:10.5281/zenodo.19464166

5. Heston TF. Resampling Fragility, Perturbation Fragility, and Why the Global Fragility Index Is Not a P-Value in Disguise. Internet Med J. 2026;1: e22059146. doi:10.5281/zenodo.22059146

6. Baer BR, Fremes SE, Gaudino M, Charlson M, Wells MT. On clinical trial fragility due to patients lost to follow up. BMC Med Res Methodol. 2021;21: 254. doi:10.1186/s12874-021-01446-z

7. Loomba R, Abdelmalek MF, Armstrong MJ, Jara M, Kjær MS, Krarup N, et al. Semaglutide 2·4 mg once weekly in patients with non-alcoholic steatohepatitis-related cirrhosis: a randomised, placebo-controlled phase 2 trial. Lancet Gastroenterol Hepatol. 2023;8: 511–522. doi:10.1016/S2468-1253(23)00068-7