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

Specification Perturbation Measures Fragility, Not Robustness: Preserving the p–fr–nb Distinction

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	12 May 2026
DOI	https://doi.org/10.5281/zenodo.20149369
Article type	Commentary
Citation	Heston TF. Minimum Specification Perturbation in Causal Inference and the Neutrality Boundary in Randomized Clinical Trial Evidence. Internet Medical Journal. 2026;1:e20149369

Abstract

The fragility index is increasingly recognized as insufficient for evaluating clinical trial evidence, and orthogonal companion metrics have been proposed across methodological traditions. A recent statistics preprint introduces Minimum Specification Perturbation as a robustness metric measuring how many analyst decisions must change to flip a confidence interval across zero. Under the Neutrality Boundary Framework's definitions, however, that construction is a fragility metric — a perturb-and-count algorithm that flips the significance classification, structurally analogous to the fragility index but in specification space rather than outcome space. The actual robustness axis is the geometric distance of the observed effect from the null parameter value, which is the role the Neutrality Boundary Framework occupies. Preserving the distinction between fragility and robustness preserves the orthogonality at the heart of the significance-fragility-robustness framework, the p-fr-nb triplet. This triplet of data analysis remains the foundation for complete statistical evidence in biomedical research.

Keywords

fragility index, minimum specification perturbation, neutrality boundary framework, statistical robustness, randomized controlled trial, evidence quality, p–fr–nb framework, clinical trial methodology

Although Minimum Specification Perturbation (MSP) has been introduced as a robustness metric capturing distance-to-falsification in causal inference (1), strictly speaking, it is a fragility metric: a count of perturbations required to flip the significance classification, structurally analogous to the fragility index but in specification space rather than outcome space. The orthogonal axis to fragility — robustness — is the geometric distance to neutrality, as described by the Neutrality Boundary Framework (NBF) (2). Distinguishing fragility from robustness is not a semantic preference; it is the conceptual move that allows two independent dimensions of statistical evidence to be reported without collapsing into a single muddled signal.

P-value-only reporting has long been understood to provide partial evidence of clinical effect (3), and a more recent editorial literature has flagged the fragility index itself as insufficient because trials are powered to minimum sample sizes that guarantee low fragility values by design (4). The biomedical literature has nonetheless used the word robustness loosely as a synonym for stability of significance classification. A trial with a low fragility index is described as fragile, and a trial with a high fragility index has been erroneously described as robust — a usage in which robustness and stability are treated as interchangeable. NBF introduced a sharper separation: fragility (fr) measures the stability of the significance classification under perturbation of the data or analyst decisions, while robustness (nb) measures the geometric distance of the observed effect from the null parameter value in standardized space (5). These are mathematically independent quantities. A trial result can be highly fragile yet far from neutrality, or highly stable yet near neutrality. Reporting one does not predict the other, which is precisely why both are required.

MSP operationalizes its distance-to-falsification construct by counting the smallest number of analyst decisions — covariate selection, estimator choice, exclusion rule — required to move a confidence interval to contain zero (1). The construction is a perturb-and-count algorithm: define a perturbation operator, apply perturbations one at a time, and record the count at which the significance classification flips. The fragility index uses the same algorithm in outcome-cell space, applying outcome flips rather than specification changes. Both metrics ask the same question — how many perturbations are needed to flip the significance classification — and under NBF both belong on the fragility axis. The perturbation operators differ; the conceptual axis does not. Calling the specification-space variant a robustness metric reverts to the pre-NBF conflation of robustness with stability and reintroduces the very ambiguity the framework was designed to resolve.

Robustness asks a different clinical question: how far is the observed treatment effect from showing no effect at all? A large, well-separated effect lies far from the no-effect line. A barely-detectable effect lies near it. Robustness puts a single number on that distance — between 0 (the result is indistinguishable from no effect) and 1 (maximal separation from no effect) — and it does so by combining the observed effect with the natural scale of variability in the data. Because every robustness value lands on the same 0-to-1 scale regardless of trial size or outcome type, robustness numbers are directly comparable across trials. That comparability makes robustness valuable for both single-trial assessment and for pooling evidence across trials in systematic reviews and meta-analyses (2). Robustness is the orthogonal-to-fragility axis in the p–fr–nb framework. MSP does not quantify robustness; it is a fragility metric.

The randomized clinical trial context reinforces the point. Pre-registration of the analytic plan — including the primary outcome, the statistical test, and the covariates included in adjustment models — bounds the analyst-decision space that MSP perturbs, making the metric near-trivial by design in prespecified randomized trials. The question that remains for a randomized clinical trial is whether the observed effect lies meaningfully far from therapeutic neutrality (robustness), and that question is answered by a geometric robustness metric, not by an additional fragility metric in a different perturbation space.

Recent meta-epidemiological evidence underscores the same point at the meta-analytic level: Cochrane pooled estimates with marginally nonsignificant p-values in the 0.05–0.20 range carry a signal of clinical relevance that legacy fragility metrics cannot adjudicate (6), and resolving that signal requires a robustness metric that quantifies geometric distance from therapeutic neutrality.

MSP is a useful contribution to the fragility family — specifically, a specification-space fragility metric well-suited to observational causal inference, where analyst-decision freedom is a substantive concern. But it is not a robustness metric, and it does not replace or duplicate the nb axis. The orthogonal robustness axis in biomedical evidence remains the geometric distance of the observed effect from therapeutic neutrality, and the p–fr–nb framework remains the complete evidence quality system for randomized clinical trials and their meta-analyses.

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. Dang H, Pham L, Nguyen M. Minimum Specification Perturbation: Robustness as Distance-to-Falsification in Causal Inference [Internet]. arXiv; 2026 [cited 2026 May 12]. Available from: https://arxiv.org/abs/2605.01579 doi:10.48550/ARXIV.2605.01579

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

3. Heston TF, King JM. Predictive power of statistical significance. World J Methodol. 2017 Dec 26;7(4):112–6. doi:10.5662/wjm.v7.i4.112

4. Cote MP, Lubowitz JH, Rossi MJ, Matzkin E. The Fragility Index Is Typically Misinterpreted and of Low Value: Clinical Trials Are Designed to Be Fragile. Arthrosc J Arthrosc Relat Surg. 2024 Nov 1;40(11):2635–7. doi:10.1016/j.arthro.2024.08.003

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

6. Sharifan A, Harrod C, Dobrescu A, Gartlehner G. Reverse fragility index shows high fragility in Cochrane meta-analyses with P values between 0.05 and 0.20: A meta-epidemiological study. J Clin Epidemiol. 2026 Apr 21;0(0). doi:10.1016/j.jclinepi.2026.112285 PubMed PMID: 42025973.