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

Reverse Fragility in Cochrane Meta-Analyses with P Values 0.05 to 0.20 Requires a Robustness Dimension

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	08 APR 2026
DOI	https://doi.org/10.5281/zenodo.19741629
Article type	Commentary
Citation	Heston TF. Reverse Fragility in Cochrane Meta-Analyses with P Values 0.05 to 0.20 Requires a Robustness Dimension. Internet Medical Journal. 2026;1:e19741629

Abstract

A meta-epidemiological analysis of 280 Cochrane meta-analyses with P-values between 0.05 and 0.20 reports high reverse fragility and interprets these findings as potentially indicating clinical relevance. That interpretation exceeds what the reverse fragility index can measure. Reverse fragility quantifies classification stability; it does not quantify distance from therapeutic neutrality. The two dimensions are orthogonal, so fragile near-null meta-analyses and underpowered detections of real effects can yield identical reverse fragility values. Completing the analysis with the neutrality boundary framework at the meta-analytic level discriminates between these cases and specifies which nonsignificant Cochrane results warrant a clinical relevance claim.

Keywords

fragility index, reverse fragility, neutrality boundary, robustness, meta-analysis, Cochrane, p-fr-nb framework, evidence synthesis

A meta-epidemiological report found that 280 Cochrane meta-analyses with P-values between 0.05 and 0.20 are reverse-fragile, which the authors imply may carry a signal of clinical relevance and challenge their treatment as settled null findings (1). However, based on the p-fr-nb framework, interpreting reverse fragility in this way overreaches what a fragility index can actually measure (2). Reverse fragility captures classification stability. It does not capture distance from therapeutic neutrality. Without that second dimension, a fragile near-null pooled estimate cannot be distinguished from a fragile underpowered detection of a real effect.

Statistical evidence has three orthogonal dimensions rather than two. Significance (p) addresses compatibility with the null hypothesis. Fragility (fr) addresses the stability of the significance classification under small perturbations. Robustness (nb) addresses the geometric distance of the observed effect from therapeutic neutrality. Only joint reporting of the p–fr–nb triplet constitutes complete statistical evidence; p alone or p with a confidence interval is partial evidence. The independence of these three dimensions is what makes the framework work: two meta-analyses can share identical fragility — both fragile or both stable — while differing sharply in their distance from neutrality, and a fragility metric alone cannot distinguish those states.

The Cochrane meta-analysis report computes reverse fragility index and reverse fragility quotient for 280 nonsignificant pooled effect estimates and observes a median reverse fragility index of 3, with P-values of 0.05 to 0.10 more fragile than those of 0.10 to 0.20 (1). That report interprets these results as suggesting that the nonsignificant findings may indicate clinical relevance and challenge their misinterpretation as robust null findings. The interpretation presumes access to the robustness dimension that reverse fragility does not provide. Two meta-analyses — one a genuine near-null pooled effect, and one an underpowered detection of a clinically meaningful effect — can produce identical reverse fragility values (2). Discriminating between them requires a metric for distance from neutrality, not another for the stability of classification or the probability under the null hypothesis. Directional refinements to fragility do not fill the gap that robustness — distance from therapeutic neutrality — fills (3).

The neutrality boundary framework provides the missing dimension for binary-outcome meta-analyses via the Risk Quotient (RQ), the standard robustness metric for independent-samples 2×2 tables (4). For any 2×2 outcome table with cell counts a, b, c, d and total N, RQ = $|ad - bc| / (N^2/4)$, with nb = RQ on the 0–1 scale and neutrality defined as independence (ad = bc). This computation requires only the aggregate event and non-event counts already published in the meta-analysis; it asks for no access to patient-level data, no simulations, and no distributional assumptions. At the meta-analytic level, complete statistical evidence is achieved by synthesizing the p–fr–nb triplet across component studies, rather than by naively summing events and non-events into a single pooled 2×2 table. RQ is the natural robustness partner to the fragility quotient family and specifically pairs with the Modified-Arm Fragility Quotient (MFQ), the allocation-fair fragility metric for 2×2 trials (5). For binary Cochrane meta-analyses, synthesizing the individual study p–fr–nb triplets constitutes

complete statistical evidence. That triplet sorts each nonsignificant meta-analysis into one of six interpretive cells defined by fragility (fragile vs. stable) and distance from independence (weak, moderate, or strong). The Sharifan cohort, with a median reverse fragility index of 3, populates the fragile category almost exclusively, and within that category, a metric for distance from neutrality resolves the decisive ambiguity: fragile and weak robustness (near therapeutic neutrality) indicates a probable true negative with an unstable verdict, while fragile and strong robustness (far from therapeutic neutrality) indicates a likely underpowered detection of a real effect. Only the latter licenses the claim that a nonsignificant Cochrane result may carry a signal of clinical relevance.

Three points support the necessity of pairing RQ with a unified fragility metric (such as MFQ) rather than relying on reverse fragility analysis. First, the classic fragility index was proposed to quantify how many outcome changes would flip a significance classification (6), and the reverse fragility extension applies the same toggling logic to nonsignificant results (7); neither metric was designed to quantify distance from neutrality, which is a geometric rather than a combinatorial property. Second, empirical validation on 129 real trials and 720,000 simulated trials confirms that fragility and robustness are orthogonal dimensions (2). An independent prior simulation of one million trials establishes provisional RQ thresholds (weak < 0.075, moderate 0.075–0.227, strong ≥ 0.227), so RQ is ready for immediate categorical interpretation, although further calibration is necessary. Third, at the Cochrane meta-analytic level, this orthogonality matters most in the subset identified as most fragile. Pooled estimates with P=0.05–0.10 sit closer to the significance boundary than those with P=0.10–0.20, which makes them more fragile (e.g., demonstrated by a median reverse fragility index of 2 versus 4 in the Sharifan cohort), but does not dictate their distance from independence. Reverse fragility is itself correlated with the P-value and therefore cannot adjudicate whether this cluster reflects fragile-but-real underpowered detections or fragile near-null pooled estimates with low precision. RQ, computed independently of the P-value from the aggregate counts, can.

Reanalysis of the 280 Cochrane nonsignificant meta-analyses by synthesizing the p–fr–nb triplets across component studies would permit classification of each result into its correct interpretive cell and identify the specific subset for which a signal of clinical relevance claim is methodologically supportable. Systematic review authors and methodologists assessing the conclusiveness of evidence should report the p–fr–nb triplet — operationalized as P-value, MFQ, and RQ for binary outcomes — rather than reverse fragility alone when inferring that a nonsignificant pooled result may indicate a clinically relevant effect.

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. 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.

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

3. Heston TF. Bidirectional fragility is a step forward but not far enough: the case for a global fragility index. Internet Med J. 2026 Apr 8;1(1):e19464166–e19464166. doi:10.5281/zenodo.19464166

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

5. Heston TF. The Modified-Arm Fragility Quotient: An Improved Metric for Assessing Robustness in Clinical Trials. Zenodo. 2025 Aug 31. doi:10.5281/zenodo.17014261

6. 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

7. Khan MS, Fonarow GC, Friede T, Lateef N, Khan SU, Anker SD, et al. Application of the reverse fragility index to statistically nonsignificant randomized clinical trial results. JAMA Netw Open. 2020 Aug 3;3(8):e2012469. doi:10.1001/jamanetworkopen.2020.12469