The Intervention Fragility Quotient: An Improved Metric for Assessing Robustness in Clinical Trials

Thomas F. Heston, MD

Clinical Assistant Professor, University of Washington School of Medicine

Clinical Associate Professor, Elson S. Floyd College of Medicine, Washington State University

ORCID: https://orcid.org/0000-0002-5655-2512

Abstract

The reproducibility crisis in biomedical research highlights the need for more reliable statistical tools beyond the p-value. Fragility metrics, such as the Fragility Index and Fragility Quotient, have been proposed to measure the robustness of clinical trial outcomes; however, both suffer from methodological limitations, including sample size dependency and imbalance between intervention and control groups. The Intervention Fragility Quotient addresses these limitations and is applicable to both equal and unequal randomizations. While the Fragility Quotient normalizes the Fragility Index to the total sample size, the Intervention Fragility Quotient contextualizes fragility within the intervention arm, thereby providing a more clinically meaningful interpretation. Worked examples illustrate the differences between these three fragility metrics. We recommend replacing the Fragility Quotient, which is distorted by unequal randomization, with the Intervention Fragility Quotient, which applies across all randomization ratios. The Intervention Fragility Quotient provides a pragmatic and interpretable tool for clinical

Preprint v1.0 (31 Aug 2025). Preferred citation: Heston TF. *The Intervention Fragility Quotient: An Improved Metric for Assessing Robustness in Clinical Trials.* Zenodo. https://doi.org/10.5281/zenodo.17014262. ORCID: https://orcid.org/0000-0002-5655-2512.

trialists, regulators, and decision-makers, with potential applications in adaptive designs, regulatory submissions, and portfolio risk assessment.

Keywords: fragility index, fragility quotient, robustness index, intervention fragility quotient, reproducibility crisis, clinical trials

Introduction

The pharmaceutical industry faces escalating challenges in drug development, characterized by high failure rates, increasing costs, and concerns about reproducibility (1–4). The traditional reliance on null hypothesis significance testing (NHST) and p-values provides a binary view of statistical inference. However, it does not adequately capture the robustness of findings (5). As a result, fragile results often advance through costly late-stage development, contributing to financial waste and delayed access to effective therapies.

Fragility metrics have been introduced to address this limitation. The Fragility Index (FI) quantifies the number of event status changes required to reverse statistical significance (6), while the Fragility Quotient (FQ) normalizes this measure by dividing the FI by the total sample size (7). Extensions such as the Percent Fragility Index (PFI), Relative Risk Index (RRI), and Robustness Index (RI) have further advanced the field of statistical fragility (8–11). However, current metrics still face key shortcomings, including dependence on total sample size and insufficient attention to group-specific effects.

Here, the Intervention Fragility Quotient (IFQ) is introduced, which normalizes fragility within the intervention arm rather than across the entire study population. This approach directly ties fragility to the treatment effect under investigation, thereby improving interpretability and clinical relevance. The IFQ can be interpreted as a revised, more accurate definition of the Fragility Quotient (FQ). In balanced designs, it approximates FQ, but in unbalanced designs, it provides a more valid representation of fragility.

Definition of Key Fragility Metrics

The intervention fragility quotient (IFQ) is related to both the fragility index (FI) and the fragility quotient. All three terms will be defined to put them in the proper context. Note that these fragility metrics are all applied to standard 2 x 2 contingency tables comparing a binary outcome between experimental and control groups. The Fisher exact test is utilized to calculate p-values.

Fragility Index (FI)

The FI is defined as the minimum number of outcome status changes (event ↔ non-event) required to reverse the statistical significance of the primary comparison. The FI restricts outcome status changes to the group with the smaller number of events. Previous authors have defined the "Reverse FI" to be the number of outcome changes to flip a non-significant finding to a significant finding. For simplicity, it is recommended to use absolute values instead. Thus, the FI is always a non-negative integer (FI ≥ 0) and reflects the absolute count of subjects that must switch outcome classification to flip the trial's statistical significance, defined as crossing a p-value threshold of 0.05. The FI is mathematically defined as:

FI = the smallest number of outcome changes in the group with fewer events needed to flip statistical significance.

Fragility Quotient (FQ)

The FQ normalizes the FI to the total sample size. The FQ is mathematically defined as:

$$FQ = \frac{FI}{N}$$

Where N = total sample size.

Intervention Fragility Quotient (IFQ)

By normalizing FI to the intervention group size instead of the total sample size, the IFQ contextualizes fragility specifically within the group with fewer events. Note that the FI only affects the group with fewer events, with no changes made to the other arm. Thus, the IFQ avoids artificial dilution by a disproportionately large control group (or vice versa) and provides a more accurate measure of the intervention's robustness. The IFQ is mathematically defined as:

$$IFQ = \frac{FI}{N(g)}$$

Where N(g) = sample size of the group to which the outcomes were changed (i.e., the arm with fewer events).

If the intervention had fewer events, then: $IFQ = \frac{FI}{N(intervention)}$

If the control had fewer events, then: $IFQ = \frac{FI}{N(control)}$

Although the FI may be computed in either the intervention or control arm (whichever has fewer events), we retain the term "Intervention Fragility Quotient" for continuity, as fragility typically arises in the intervention arm in most clinically relevant scenarios.

The IFQ differs from the FQ, which uses the total sample size in the denominator, and from the FI, which provides an absolute count without normalization. The IFQ, therefore, isolates

fragility to the group where outcomes were altered and is applicable across both equal and unequal randomization ratios.

Illustrative Examples

Example A: Balanced 1:1 randomization (observed data)

The first example looks at a study with equal randomization (Table 1A).

Table 1A. Example A (50/50 randomization, observed data).

Group	Event	Non-event	Total
Experimental	20	30	50
Control	30	20	50
Total	50	50	100

p = 0.0713

Calculation steps:

1. Apply Fisher exact test → p = 0.0713
2. Identify arm with fewer events
3. Flip outcomes until significance reversed → p 0.0449
4. FI = 1, FQ = 0.01, and IFQ = 0.02 (Table 1B).

Table 1B. Example A (50/50 randomization, after FI applied)

Group	Event	Non-event	Total
Experimental	19	31	50

Control	30	20	50
Total	50	50	100

p = 0.0449

This indicates that in balanced designs, IFQ provides the same information as FQ and equals the FI multiplied by 2, thereby preserving interpretability.

Example B: Unbalanced randomization (80 intervention / 20 control)

This second example examines the consequences of unequal study allocation (Table 3).

Table 2A. Example B (80/20 randomization, observed data)

Group	Event	Non-event	Total
Experimental	50	30	80
Control	9	11	20
Total	58	42	100

p = 0.2047

Calculation steps:

1. Apply Fisher exact test → p = 0.2047 (non-significant)
2. Identify arm with fewer events
3. Flip outcomes until significance reversed → p = 0.0418
4. FI = 2, FQ = 0.02, and IFQ = 0.10, a five-fold difference (Table 2B).

Table 2B. Example B (80/20 randomization, after FI applied)

Group	Event	Non-event	Total
Experimental	50	30	50

Control	9	13	50
Total	59	43	100

p = 0.0419

These examples show that for equal randomization, IFQ = FQ * 2 in all cases. However, in unequal randomization, the IFQ will be significantly different than the FQ. In such cases, the IFQ provides a more accurate picture of the magnitude of changes made to arm with fewer events.

Applicability and Advantages of the IFQ

The advantages of the IFQ become clear when considering both balanced and unbalanced trial designs. In a study with equal allocation, the IFQ equals twice the value of the FQ. In this situation, the IFQ contains all of the information in the FQ. It preserves interpretability and consistency with existing fragility measures.

The strength of the IFQ emerges most clearly in trials with unequal randomization, such as 60/40 or 80/20 designs. In these situations, the FQ can obscure fragility by spreading the denominator across both study arms. The IFQ avoids this dilution by anchoring interpretation to the arm where fragility arises, providing a more clinically relevant picture of how easily results could be overturned.

Importantly, the IFQ is not limited to one type of design. Because it normalizes fragility within the group that determines statistical reversal, it remains valid across both equal and unequal randomizations. This universality extends to adaptive and pragmatic designs as well, making the IFQ a flexible tool for contemporary clinical research.

By reframing fragility in terms of the group that drives significance reversal, the IFQ delivers clearer insights into the stability of treatment effects. This added interpretability

can support better decisions in trial design, regulatory review, and even portfolio risk assessment.

Discussion

The IFQ builds upon earlier fragility metrics by offering a measure that is both more clinically relevant and more broadly applicable. Unlike the FQ, which can distort fragility in the setting of unequal allocation, the IFQ normalizes fragility within the arm that determines statistical reversal. In doing so, it focuses attention on the stability of the treatment effect itself rather than diluting interpretation across the entire study sample (6,7).

Expressing fragility as a proportion also improves interpretability. The IFQ permits meaningful comparison across trials of different sizes and randomization ratios, thereby complementing traditional effect size estimates and confidence intervals. This feature is particularly important in modern clinical research, where unequal or adaptive randomizations are increasingly common (12–14).

The implications extend beyond methodology. In pharmaceutical development, fragile results that appear statistically persuasive can lead to costly late-stage failures. By quantifying how robust—or brittle—trial conclusions are, the IFQ provides a framework that aligns with decision-making in drug development and regulatory review. It may therefore contribute to more accurate go/no-go determinations and improved portfolio-level risk assessment (3,4).

Like all fragility measures, the IFQ has limitations. At present it is limited to binary outcomes analyzed in 2×2 contingency tables. Its role in adaptive, non-inferiority, and equivalence designs remains to be clarified, and future work should extend the concept to continuous and time-to-event outcomes. Another promising direction is integration with

Bayesian and adaptive statistical frameworks, which are becoming increasingly influential in trial design.

Conclusion

The IFQ advances fragility analysis by situating robustness within the arm where significance is most vulnerable, rather than across the entire study population. This enhances interpretability, strengthens clinical relevance, and improves decision-making utility. Because it expresses fragility as a proportion of the arm in which the FI is applied, IFQ allows fair comparison across sample sizes and randomization ratios.

Crucially, FQ should be retired. Dividing FI by the total sample size adds no value in balanced trials and becomes misleading under unequal allocation; IFQ resolves this by normalizing to the relevant arm. We recommend reporting FI and IFQ together in clinical trial results and discontinuing use of FQ to improve transparency and the appraisal of result stability.

Bibliography

1. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005 Aug 30;2(8):e124. DOI: 10.1371/journal.pmed.0020124. PMID: 16060722. PMCID: PMC1182327.
2. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012 Mar 28;483(7391):531–3. DOI: 10.1038/483531a. PMID: 22460880.
3. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016 May;47:20-33. DOI: 10.1016/j.jhealeco.2016.01.012. PMID: 26928437.

4. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA. 2020 Mar 3;323(9):844-853. DOI: 10.1001/jama.2020.1166. PMID: 32125404. PMCID: PMC7054832.

5. Wasserstein RL, Lazar NA. The ASA Statement on *p* -Values: Context, Process, and Purpose. Am Stat. 2016 Apr 2;70(2):129–33. DOI: 10.1080/00031305.2016.1154108.

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):622-628. DOI: 10.1016/j.jclinepi.2013.10.019. PMID: 24508144.

7. Ahmed W, Fowler RA, McCredie VA. Does sample size matter when interpreting the fragility index? Crit Care Med. 2016 Nov;44(11):e1142-e1143. DOI: 10.1097/CCM.0000000000001976.

8. Heston TF. The Percent Fragility Index. International Journal of Scientific Research. 2023 Jul;12(7):9–10. DOI: 10.5281/zenodo.8215815.

9. Heston TF. The robustness index: going beyond statistical significance by quantifying fragility. Cureus. 2023 Aug 30;15(8):e44397. DOI: 10.7759/cureus.44397. PMID: 37791215. PMCID: PMC10542213.

10. Heston TF. Statistical Significance Versus Clinical Relevance: A Head-to-Head Comparison of the Fragility Index and Relative Risk Index. Cureus. 2023 Oct 26;15(10):e47741. DOI: 10.7759/cureus.47741. PMID: 37899890. PMCID: PMC10602368.

11. Heston TF. Redefining significance: robustness and percent fragility indices in biomedical research. Stats. 2024 Jun 17;7(2):537-548. DOI: 10.3390/stats7020033.

12. Park JJ, Thorlund K, Mills EJ. Critical concepts in adaptive clinical trials. Clin Epidemiol. 2018 Mar 23;10:343–51. DOI: 10.2147/CLEP.S156708. PMID: 29606891. PMCID: PMC5868584.

13. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic

factors in the controlled clinical trial. Biometrics. 1975 Mar;31(1):103–15. DOI: 10.2307/2529712. PMID: 1100130.

14. Chow S-C, Chang M. Adaptive design methods in clinical trials. 2nd ed. Boca Raton; 2011.