

The Application of Reverse Fragility Index to Plastic Surgery Randomized Controlled Trials with Statistically Nonsignificant Outcomes

Rodney Ahdoot, BS, Trista M. Benitez, MD, MPH, Chien-Wei Wang, MD, Kevin C. Chung, MD, MS

Section of Plastic Surgery, Department of Surgery, University of Michigan Medical School, Ann Arbor, MI



BACKGROUND

Randomized controlled trials (RCTs) use p-values to assess significance, often relying on a threshold of 0.05. However, important study factors such as sample size, statistical power, and design quality may be overlooked. The fragility index (FI) measures how easily significant results can become nonsignificant, while the reverse fragility index (RFI) does the same for nonsignificant results. The reverse fragility quotient (RFQ) adjusts RFI for sample size. These metrics are particularly useful in plastic surgery, where small sample sizes and low event rates raise the risk of false-negative findings. In this study, we applied RFI and RFQ to plastic surgery RCTs with nonsignificant outcomes and explored factors associated with study robustness.

METHODS

A systematic literature search of the 10 highest impact journals covering plastic and reconstructive surgery topics was conducted to identify RCTs published between 2000 to 2023. Eligibility criteria: (1) human participants, (2) used a 1:1 parallel arm study design, (3) published in English, and (4) contained at least one statistically nonsignificant primary dichotomous outcome mentioned in the abstract ($p \geq 0.05$ or a 95% confidence interval containing the null value).

The RFI was calculated by modifying the outcome events in a 2 x 2 contingency table. This was done by first identifying the group (e.g. intervention vs. control) with the fewer number of events, then iteratively subtracting events while simultaneously adding nonevents to this group, maintaining the number participants to achieve a p-value less than 0.05 by Fisher’s exact 2-sided test. The total number of event reversals required to achieve significance was recorded as the RFI.

The RFQ for each outcome was calculated by dividing the RFI by the sample size. Bivariate analysis was performed to investigate the association of trial characteristics with the RFI.

Journal	h5-index (2023)
Plastic and Reconstructive Surgery	61
Aesthetic Surgery Journal	45
Plastic and Reconstructive Surgery Global Open	40
Journal of Plastic, Reconstructive & Aesthetic Surgery	39
Aesthetic Plastic Surgery	36
Annals of Plastic Surgery	32
Journal of Craniofacial Surgery	31
Clinics in Plastic Surgery	30
JAMA Facial Plastic Surgery	29
Microsurgery	28

Table 1. Top 10 Plastic Surgery Journals by H5-Index (2023)

2 x 2 Contingency Table			
	Event	Nonevent	Total
Group 1	a	b	a + b
Group 2	c	d	c + d

Initial (p = 0.212)			
	Event	Nonevent	Total
Group 1	5	19	24
Group 2	10	14	24

1 st iteration (p = 0.111)			
	Event	Nonevent	Total
Group 1	4	20	24
Group 2	10	14	24

2 nd iteration (p = 0.049)			
	Event	Nonevent	Total
Group 1	3	21	24
Group 2	10	14	24

Figure 1. Sample RFI calculation with 2x2 contingency table

RESULTS

Study characteristics	n (%)
Plastic surgery area of RCT	
Breast	12 (30)
Cosmetic	12 (30)
Craniofacial	9 (22.5)
Reconstructive	4 (10)
Hand/Upper Extremity	2 (5)
Experimental	1 (2.5)
Region of Origin	
Asia	11 (27.5)
Europe	15 (37.5)
Others	9 (22.5)
United States	5 (12.5)
Enrollment Type	
Single Center	39 (97.5)
Multicenter	1 (2.5)
Blinding	
Double Blinding	10 (25)
Single Blinding	7 (17.5)
No Blinding	12 (31)
Not Reported	11 (27.5)
Loss to Follow-Up	
Reported	28 (70)
Not Reported	10 (25)
Unclear	2 (5)
Power Analysis	
Yes	19 (47.5)
No	21 (52.5)
Mention of Possible Underpowering	
Yes	14 (35)
No	26 (65)

Table 2. Characteristics of included RCT studies.

Outcome characteristics	Median (IQR)
RFI	4 (3 – 5)
RFQ	0.053 (0.038 – 0.080)
Total sample size	76 (49 – 129)
Total events	10 (3 – 19)
Total loss to follow-up (n=51)*	0 (0 – 6.5)
Loss to follow-up compared to RFI (n=51)*	
Loss to follow-up > RFI, n (%)	16 (31.4)
Loss to follow-up < RFI, n (%)	35 (68.6)

Table 3. Characteristics of analyzed nonsignificant outcomes.

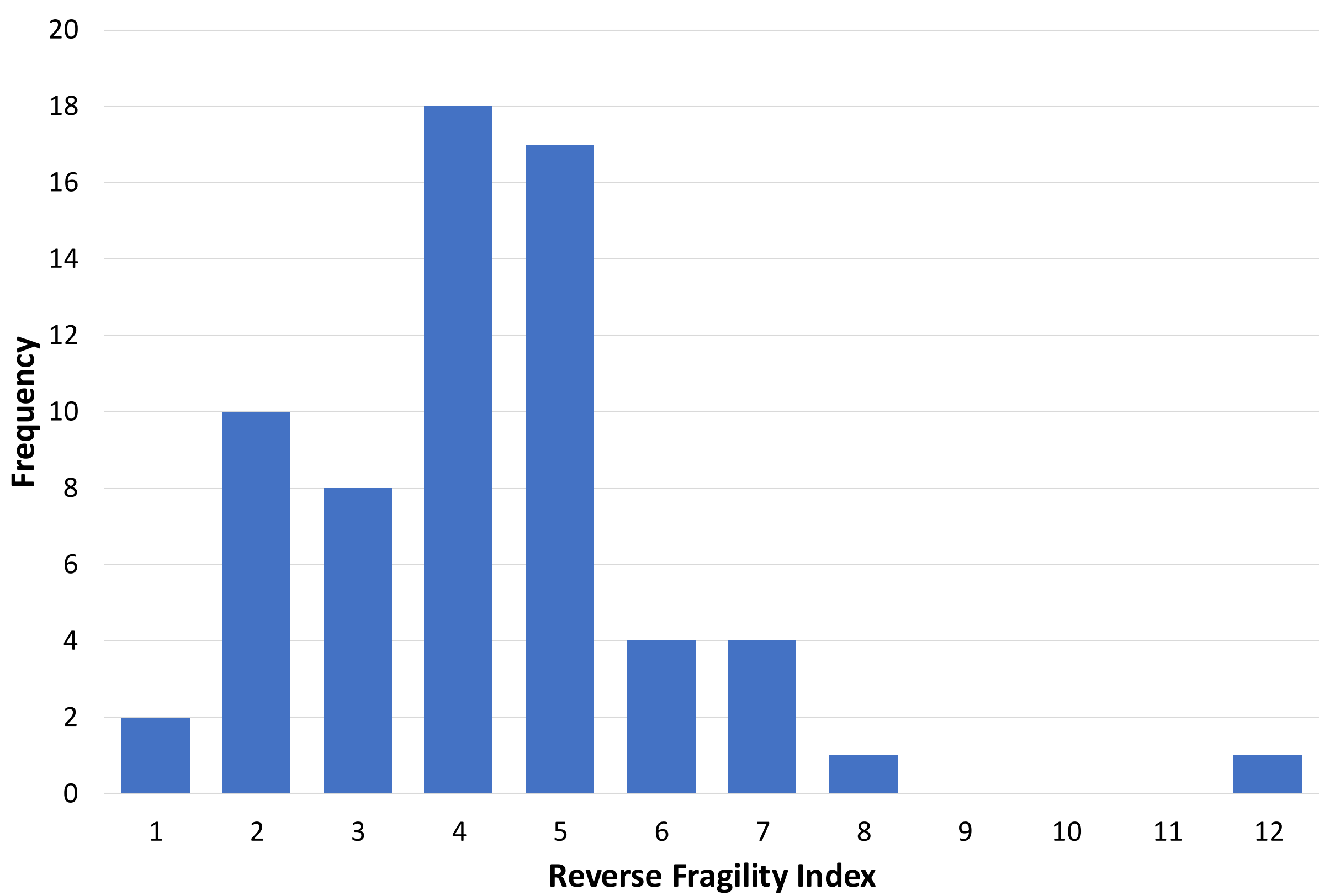


Figure 2. Distribution of reverse fragility index across 65 study outcomes.

We screened 1,140 articles and included 40 RCTs with 65 statistically nonsignificant, dichotomous primary outcomes. Most studies were published in Plastic & Reconstructive Surgery (35%).

Many trials lacked key methodological reporting: 30% did not clearly report loss to follow-up, 53% did not report a power analysis, and 35% noted potential underpowering.

The median RFI was 4 (IQR 3–5); RFQ was 0.053 (IQR 0.038–0.08). The median sample size was 76.

Of the 51 outcomes with clear loss-to-follow-up data, 31% had more loss than the corresponding RFI.

Studies that acknowledged underpowering had significantly higher RFIs than those that did not (5 vs 4; $p = 0.024$). No other significant associations were found.

CONCLUSIONS

- The findings of RCTs in the plastic surgery literature demonstrate fragility as the preservation of nonsignificant results rested on a small number of events.
- Incorporating the RFI into reporting standards can enhance the interpretation of study results by showing the number of event reversals that are needed to achieve significance. We recommend the routine reporting of the RFI and loss to follow-up alongside p-values, especially in studies with limited sample sizes.

QUESTIONS

rahdoot@med.umich.edu