Statistically Fragile Evidence in Vertebroplasty Trials for Vertebral Compression Fractures: A Systematic Review

Michael Li, Alexander Yu, Kareem S Mohamed, Mark Ambrose Kurapatti, Junho Song, Jonathan J Huang, Prabhjot Singh, Yazan Alasadi, Abhijeet Grewal, Avanish Yendluri, Charlene W Cai, Nikan Namiri, John Joseph Corvi, Jun Sup Kim, Samuel Kang-Wook Cho

INTRODUCTION:

Randomized clinical trials (RCTs) on Vertebroplasty are crucial for guiding the treatment of vertebral compression fractures (VCFs), but their overlooked statistical fragility can undermine clinical reliability. Minor outcome changes may overturn significant findings, risking unreliable evidence and impacting patient care. This study assessed the fragility of significant outcomes in Vertebroplasty RCTs, hypothesizing high sensitivity to such changes.

METHODS:

PubMed, Embase, and MEDLINE were searched for RCTs on Vertebroplasty reporting dichotomous outcomes. The Fragility Index (FI) and reverse Fragility Index (rFI) quantified the number of outcome reversals needed to change statistical significance for significant and nonsignificant results, respectively. The Fragility Quotient (FQ) was calculated as the FI divided by the study sample size. Subgroup analysis was conducted by outcome category.

RESULTS:

A total of 276 outcomes from RCTs were analyzed. The median Fragility Index (FI) was 5 (IQR: 4–5), with a FQ of 0.053 (IQR: 0.019–0.088). Statistically significant outcomes (n=36) had a median FI of 3 (IQR: 2–4) and FQ of 0.034 (IQR: 0.018–0.051), while nonsignificant outcomes (n=240) showed a median FI of 5 (IQR: 4–5) and FQ of 0.062 (IQR: 0.021–0.088). Fracture-related outcomes were the most robust (FI: 5, FQ: 0.088), while cement leakage was the most fragile (FI: 3, FQ: 0.041). Pain outcomes had an FI of 5 (FQ: 0.062), and complications and vertebroplasty vs. kyphoplasty outcomes were more robust (FI: 5, FQ: 0.013). Patients lost to follow-up exceeded the FI in 79% of outcomes.

DISCUSSION AND CONCLUSION:

The statistical findings in Vertebroplasty RCTs are fragile and warrant cautious interpretation. A small number of outcome reversals or consistent postoperative follow-up can shift the significance of the results. Standardized reporting of P values alongside FI and FQ metrics is recommended to help clinicians evaluate the robustness of study findings.

Figure 1. Demonstration of statistical significance reversal using a 2 x 2 contingency table with a resulting fragility index (FI) = 3. *P*-values were calculated using a two-tailed Fisher exact test.

	(+)	(-)
Treatment A	14	72
Treatment B	3	83
		0.009

	(+)	(-)
Treatment A	14	72
Treatment B	6	80
		0.094

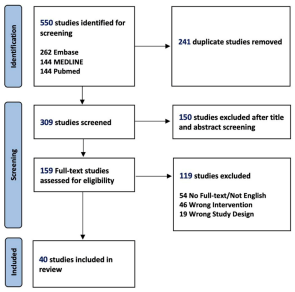

Table 2: Statistical Fragility of Overall Outcomes

	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)	# of Patients Lost to Follow Up > FI (%)
All RCT Outcomes	276	5(4-5)	0.053(0.019-0.088)	79.35%
Significant Outcomes (P<0.05)	36	3(2-4)	0.034(0.018-0.051)	55.56
Nonsignificant Outcomes (P≥0.05)	240	5(4-5)	0.062(0.021-0.088)	80.0

Table 3: Statistical Fragility of Subgroup Outcomes

	Number of outcomes	FI, Median (IQR)	FQ, Median (IQR)
Complications	50	4(4-5)	0.013(0.010-0.036)
Pain	76	5(4-7)	0.062(0.035-0.088)
Cement Leakage	25	3(2-4)	0.041(0.026-0.063)
Fracture	93	5(4-5)	0.088(0.047-0.088)
Vertebroplasty vs Kyphoplasty	65	5(4-5)	0.013(.010-.029)