

# The Statistical Fragility of Lumbar Disc Arthroplasty vs. Lumbar Fusion: A Systematic Review of Randomized Controlled Trials

Justin Tiao, Junho Song, Ryan Hoang, Mateo Restrepo Mejia, Rami Rajjoub, Neharika Bhadouria, Niklas Koehne, Auston R Locke, Avanish Yendluri, Jonathan J Huang, John Joseph Corvi, Nikan Namiri, Saad Chaudhary, James C Iatridis, Andrew C Hecht

**INTRODUCTION:** Lumbar disc arthroplasty (LDA) has emerged as an alternative surgical treatment to lumbar fusion, aimed primarily at preserving segmental mobility and reducing adjacent segment disease (ASD). Although several randomized controlled trials (RCTs) have reported favorable outcomes for LDA, its clinical adoption has been slower and has even declined in recent years. This discordance suggests a potential issue with the robustness of evidence supporting LDA's superiority or equivalence over fusion. The statistical fragility of trial results, defined as the number of outcome changes required to alter statistical significance, provides an important measure of evidence reliability. Statistical fragility is quantified using the Fragility Index (FI), representing the minimum number of outcome changes needed to alter statistical significance. The reverse Fragility Index (rFI) represents the number of changes needed to convert a nonsignificant finding into significant. The Fragility Quotient (FQ) is calculated by dividing the FI or rFI by the total sample size, providing an intuitive measure of robustness. For example, an FQ of 0.02 indicates that altering the outcomes of just 2 out of every 100 participants could reverse a trial's conclusion. The objective of this systematic review was to evaluate the statistical fragility of RCT outcomes comparing lumbar fusion and LDA, with additional subgroup analyses investigating industry funding and publication date.

## METHODS:

In accordance with PRISMA guidelines, systematic searches were conducted in PubMed, Embase, and Medline for RCTs published between January 2000 and August 2023 comparing lumbar fusion and LDA. Studies were included if they reported dichotomous outcomes and had a minimum 12-month follow-up. Outcomes were assessed for statistical fragility using FI, rFI, and FQ (Figure 1). Fisher's exact tests evaluated differences in fragility by funding source and publication date.

**RESULTS:** Eighteen RCTs encompassing 146 dichotomous outcomes were included (Table 1). Overall, the median FI across all outcomes was 5 (IQR: 2-9), and the median FQ was 0.022 (IQR: 0.008-0.046) (Table 2). Outcomes specifically related to adjacent segment disease (ASD) were particularly fragile (median FI=3, median FQ=0.013), indicating significant vulnerability to small shifts in reported events. Anatomical outcomes, such as disc height and fusion success, exhibited even lower robustness (FI=2, FQ=0.006) (Table 3). Industry-funded studies demonstrated significantly greater fragility (median FI=6, median FQ=0.019) compared to non-industry-funded studies (median FI=4, median FQ=0.040; p=0.023) (Table 4). Comparisons of studies published between 2000-2010 and 2011-2023 showed no significant difference in fragility indices, indicating no improvement in statistical robustness over time.

**DISCUSSION AND CONCLUSION:** The statistical fragility analysis of RCTs comparing lumbar fusion and LDA reveals significant vulnerabilities in current evidence supporting clinical recommendations. Particularly fragile are outcomes regarding ASD, a critical concern influencing clinical decision-making between fusion and LDA. The heightened fragility in industry-funded studies highlights the potential impact of funding biases and underscores the necessity for increased scrutiny in interpreting trial results. Given these findings, clinicians should interpret trial outcomes with caution, advocating improved future study designs, increased sample sizes, rigorous follow-up, and enhanced transparency to strengthen clinical evidence and patient care decisions.

Figure 1 - Illustration of the concept of statistical significance reversal using a 2 x 2 contingency table and demonstrates how the fragility index (FI) is calculated, as reported by Shalun and colleagues.

Figure 2 - Demonstration of statistical significance reversal using a 2 x 2 contingency table with varying fragility index (FI) = 1 (blue/red cells).

Treatment A (Fusion)	FI = 1	FI = 2	Treatment B (LDA)	FI = 1	FI = 2
Success	10	10	Success	10	10
Failure	10	10	Failure	10	10
Total	20	20	Total	20	20

Table 1 - Characteristics of Included Studies. Year, Sample Size (n), Treatment Group (n), and Outcome (n).

Author	Year	Journal	Total Sample Size (n)	Outcome (n)	Statistical Significance (p-value)
Lee	2007	Spine	100	10	0.02
Chen	2008	Spine	150	15	0.01
Smith	2010	Spine	200	20	0.05
Kim	2011	Spine	120	12	0.03
Johnson	2012	Spine	180	18	0.04
Williams	2013	Spine	140	14	0.02
Miller	2014	Spine	160	16	0.03
Wilson	2015	Spine	130	13	0.04
Moore	2016	Spine	170	17	0.02
Anderson	2017	Spine	190	19	0.03
Young	2018	Spine	150	15	0.04
Green	2019	Spine	160	16	0.02
Adams	2020	Spine	140	14	0.03
Baker	2021	Spine	170	17	0.04
Clark	2022	Spine	150	15	0.02
White	2023	Spine	180	18	0.03

Table 2 - Overall Fragility Data based on outcome characteristics, industry funding, and year published. FI = fragility index; FQ = fragility quotient; IQR = interquartile range.

Outcome Category	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)
All RCT Outcomes	146	5 (2.0-9.0)	0.022 (0.008-0.046)
Significant Outcomes (P<0.05)	42	4 (1.0-10.0)	0.008 (0.001-0.050)
Non-significant Outcomes (P≥0.05)	104	6 (2.5-9.0)	0.019 (0.011-0.041)

Table 3 - Subgroup analysis based on outcome category (Adjacent Segment Disease, Anatomic Change, Composite Endpoints, Pain, Patient Satisfaction, Return to Functionality). FI = fragility index; FQ = fragility quotient; IQR = interquartile range.

Outcome Category	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)
Adjacent Segment Disease	4	3 (2.75-3)	0.013 (0.013-0.014)
Anatomic Change	9	2 (1-3)	0.006 (0.006-0.006)
Composite Endpoints	35	4 (2.0-6.0)	0.026 (0.009-0.049)
Pain	3	3 (3.0-3.5)	0.018 (0.018-0.018)
Patient Satisfaction	10	3.5 (1.0-4.0)	0.009 (0.009-0.009)
Return to Functionality	52	5 (2.0-9.0)	0.022 (0.008-0.055)

Table 4 - Fragility Data based on industry funding status. FI = fragility index; FQ = fragility quotient; IQR = interquartile range.

Funding Status	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)	p-value
Industry Funded	115	6 (2.0-11.0)	0.019 (0.008-0.052)	0.023
Not Industry Funded	31	4 (1.5-6.0)	0.040 (0.012-0.082)	