The Fragility of Statistical Findings in Cervical Disc Arthroplasty: A Systematic Review of Randomized Controlled Trials

Michael N. Megafu[1], Emmanuel Megafu, Janet Tuong Vi Nguyen, Elisabeth Point Du Jour, Wesley Bronson[2], James D Lin[3], Andrew C Hecht[4], Robert L. Parisien[3]
[1]Department of Orthopaedic Surgery, University of Connecticut, [2]Mount Sinai Hospital, [3]Mount Sinai Health System, [4]Mt Sinai Med Ctr & Schl of Med

INTRODUCTION: This study employs both the fragility index (FI) and fragility quotient (FQ) to assess the level of robustness in the cervical disc arthroplasty (CDA) literature. We hypothesize that dichotomous outcomes involving CDA would exhibit statistical vulnerability.

METHODS:

A PubMed search evaluated dichotomous data for randomized controlled trials (RCTs) in CDA literature from 2000 to 2023. Each outcome's FI was calculated by reversing a single outcome event until significance was reversed. The FQ was calculated by dividing each fragility index by the study sample size. The interquartile range (IQR) was also calculated for the FI and FQ.

RESULTS: Of the 1561 articles screened, 111 met the search criteria, with 35 RCTs evaluating CDA included for analysis. Six hundred and ninety-three outcome events with 130 significant ($P < 0.05$) outcomes and 563 nonsignificant ($P \geq 0.05$) outcomes were identified. The overall FI and FQ for all 693 outcomes were 5 (IQR 3-7) and 0.019 (IQR 0.011-0.043). Fragility analysis of statistically significant and nonsignificant outcomes revealed an FI of 5. All of the studies reported loss to follow-up (LTF) data where 65.7% (23) did not report or reported an LTF greater or equal to 5.

DISCUSSION AND CONCLUSION: The literature regarding CDA RCTs lacks statistical robustness and may misrepresent the conclusions solely using the *P* value. By implementing the FI and FQ along with the *P* value, the interpretation and contextualization of the clinical data surrounding CDA will be better understood.

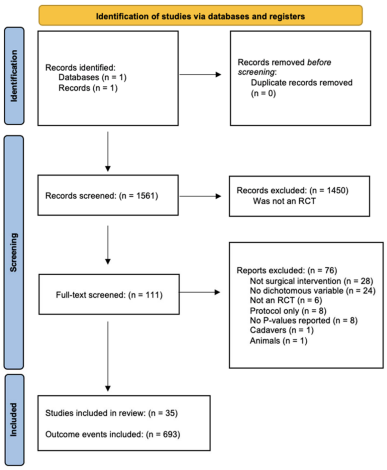

Table 1: Demonstration of Reversal Significance with a Fragility of 1

	Outcome A	Outcome B	*P* Value
Scenario 1			
Treatment A	13	27	
Treatment B	5	35	0.059
Scenario 2			
Treatment A	14	26	
Treatment B	5	35	0.034

Table 2: Overall Fragility Data and Analysis of Subgroups[a]

Characteristic	Events	Fragility Index (IQR)	Fragility Quotient (IQR)
All trials	693	5 (3-7)	0.019 (0.011-0.043)
Outcome significance[b]			
P < 0.05	130	5 (2-11)	0.017 (0.006-0.044)
P > 0.05	563	5 (4-7)	0.019 (0.011-0.043)
Comparing outcome FI to LTF[c]			
FI < LTF	565	5 (3-7)	0.016 (0.009-0.033)
FI > LTF	128	5 (4-7)	0.055 (0.025-0.092)
Year of publication			
2000 – 2007	43	5 (4-5)	0.052 (0.043-0.073)
2008 – 2015	434	5 (3-7)	0.016 (0.009-0.036)
2016 – 2023	216	5 (3-10)	0.021 (0.011-0.044)
Journals			
Clinical Orthopaedics and Related Research	19	5 (4-7)	0.060 (0.048-0.084)
Clinical Spine Surgery	26	4.5 (4-8)	0.059 (0.040-0.133)
International Orthopaedics	20	4 (4-6)	0.054 (0.045-0.062)
Journal of Bone and Joint Surgery	62	5 (3-6)	0.017 (0.011-0.030)
Journal of Neurosurgery: Spine	156	6 (3-10)	0.025 (0.011-0.045)
Spine (Phila Pa 1976)	246	5 (3-7)	0.014 (0.009-0.024)
Journal of the Pakistan Medical Association	6	4.5 (4-7)	0.107 (0.095-0.167)
Journal of Spinal Disorders and Techniques	42	5 (5-9)	0.043 (0.011-0.067)
Neurosurgical Focus	18	4 (4-5)	0.073 (0.073-0.091)
Orthopedics	5	4 (4-5)	0.167 (0.167-0.208)
The Spine Journal	93	5 (3-8.5)	0.014 (0.008-0.028)

[a]FI, fragility index; IQR, interquartile range; LTF, lost to follow-up.
[b]$P \leq 0.05$ represents the significant outcome subgroup and $P > 0.05$ represents the insignificant outcome subgroup
[c]FI ≤ LTF represents the outcome subgroup where the FI was less than the number of patients LTF. FI > LTTF represents the outcome subgroup where the FI was greater than the number of patients LTF.