The Statistical Fragility of Collagenase Treatment For Dupuytren's Contracture: A Systematic Review of Randomized Controlled Trials

Auston R Locke, Niklas Koehne, Sri Guttikonda, Charlene Cai, Christoph Alexander Schroen, Jamie Kator, Jaehon M Kim, Michael R Hausman

INTRODUCTION:

Randomized controlled trials (RCTs) are widely regarded as the gold standard for generating high-quality evidence and informing both surgical and nonsurgical clinical decision-making. In the context of Dupuytren's contractures (DCs), a condition with variable presentation and an evolving treatment landscape that includes surgical fasciectomy, percutaneous needle aponeurotomy, and injectable collagenase, RCTs play a critical role in guiding evidence-based care. Despite this, the statistical robustness of reported outcomes in these trials is rarely examined. Strict reliance on p-values alone has come under increasing scrutiny, as statistically significant findings may not always be clinically reliable. The concept of statistical fragility, characterized by the fragility index (FI) and fragility quotient (FQ) provides a complementary tool to evaluate the stability of dichotomous outcomes in clinical trials. This study aims to systematically assess the statistical fragility of RCTs evaluating DC treatment and to identify which categories of outcomes, especially those relating to treatment failure, are most vulnerable to outcome reversal. We hypothesized that DC RCTs would exhibit high statistical fragility, particularly in adverse event and failure-related outcomes.

METHODS:

We performed a systematic review in accordance with PRISMA guidelines. PubMed, Embase, and MEDLINE were queried for English-language RCTs published between January 1, 2000, and September 30, 2024, assessing any interventions for DC treatment. Studies were included if they reported dichotomous, categorical outcomes related to clinical efficacy, adverse events, or treatment success/failure. Studies without accessible full-text, non-English articles, and non-RCTs were excluded. For each outcome, we calculated the FI (for significant results) or reverse FI (rFI, for nonsignificant results), representing the number of event reversals needed to change the statistical significance. FQ was calculated by dividing the FI or rFI by the sample size for each respective outcome. Additionally, we categorized outcomes into clinical themes (e.g., adverse events, treatment success) and stratified studies by publication era (2004-2015 vs. 2016-2024). We also performed a secondary analysis excluding outcomes where the number of patients lost to follow-up (LTF) exceeded the corresponding FI or rFI.

RESULTS:

A total of 179 RCTs were screened, of which 19 met inclusion criteria, yielding 173 dichotomous outcomes for analysis. The median FI across all outcomes was 4 (IQR 2-6), and the median FQ was 0.075 (IQR 0.04-0.139), indicating that in half of the outcomes, a reversal of the event status of 4 patients, or approximately 7.5% of the cohort, would alter the statistical significance. Statistically significant outcomes (median FQ = 0.095) were more fragile than those that were statistically nonsignificant (median FQ = 0.063), suggesting that even nominally "positive" findings are susceptible to minor shifts in data. Of note, 15.1% of all outcomes (27/179) had an LTF count exceeding the respective FI, raising concerns about potential data instability. When stratified by category, adverse events (82 outcomes), treatment success (62 outcomes), and treatment failure (26 outcomes) were the most commonly reported endpoints, with median FQs of 0.055, 0.11, and 0.62, respectively. Treatment failure had the highest fragility, implying a particular need for caution when interpreting results in this domain.

DISCUSSION AND CONCLUSION: This is the first known study to systematically evaluate the statistical fragility of RCTs focused on Dupuytren's contracture treatments. Our findings demonstrate that outcomes reported in these trials are often statistically fragile, particularly those related to treatment failure and adverse events. Given the low FI and FQ values, many trial outcomes may be more vulnerable to change than their p-values imply. While LTF did not directly alter the majority of outcomes in this analysis, its presence above the FI threshold in over 15% of cases emphasizes the importance of complete follow-up in fragility analyses. We recommend that future RCTs in this field report fragility indices alongside traditional significance testing to better inform clinicians and enhance the interpretability of trial findings. Incorporating fragility metrics could help mitigate overreliance on p-values, providing a more nuanced understanding of the robustness and clinical applicability of RCT findings in Dupuytren's contracture management.

Figure 1 - Demonstration of statistical significance reversal using a 2 x 2 contingency table with a resulting fragility index (FI) = 1 (Bedaiwaa et al.)

	Clinical Success (+)	No Clinical Success (-)
Treatment A (Collagenase)	8	2
Treatment B (Placebo)	3	7
		p = 0.031

	Clinical Success (+)	No Clinical Success (-)
Treatment A (Collagenase)	7	3
Treatment B (Placebo)	3	7
		p = 0.080

Table 2 – Subgroup analysis based on outcome category

	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)
Adverse Event	82	4 (2-6)	0.055 (0.028-0.12)
Treatment Failure	26	5 (4-6)	0.062 (0.041-0.092)
Treatment Success	62	4.5 (3-8)	0.11 (0.064-0.18)
Patient Satisfaction	4	5 (4.75-5)	0.076 (0.071-0.078)

Table 1 – Fragility data based on year published (2004-2015, 2016-2024)

	Number of Outcomes	FI, Median (IQR)	FQ, Median (IQR)
Published 2004-2015	107	4 (1.5-9)	0.076 (0.029-0.14)
Published 2016-2024	67	5 (4-6)	0.06 (0.043-0.10)