Title

Fragility of overactive bladder medication clinical trials: A systematic review

Journal

Neurourology and Urodynamics, 43(7)

ISSN

0733-2467

Authors

Li, Kevin D
Venishetty, Nikit
Fernandez, Adrian M
et al.

Publication Date

2024-04-09

DOI

10.1002/nau.25468

Received: 26 March 2024 | Accepted: 1 April 2024

DOI: 10.1002/nau.25468

REVIEW

Neurourology Urodynamics | SUFU | WILEY

Fragility of overactive bladder medication clinical trials: A systematic review

Kevin D. Li[1,2] | **Nikit Venishetty**[3] | **Adrian M. Fernandez**[1] | **Nizar Hakam**[1] | **Umar Ghaffar**[1] | **Shiv Gupta**[1] | **Hiren V. Patel**[1] | **Benjamin N. Breyer**[1,2]

[1]Department of Urology, University of California San Francisco, San Francisco, California, USA

[2]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

[3]Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso, Texas, USA

Correspondence
Kevin D. Li, Department of Urology, University of California San Francisco, San Francisco, CA, USA.
Email: kevin.d.li@ucsf.edu

Abstract

Purpose: Overactive bladder (OAB) syndrome significantly impairs quality of life, often necessitating pharmacological interventions with associated risks. The fragility of OAB trial outcomes, as measured by the fragility index (FI: smallest number of event changes to reverse statistical significance) and quotient (FQ: FI divided by total sample size expressed as a percentage), is critical yet unstudied.

Materials and Methods: We conducted a systematic search for randomized controlled trials on OAB medications published between January 2000 and August 2023. Inclusion criteria were trials with two parallel arms reporting binary outcomes related to OAB medications. We extracted trial details, outcomes, and statistical tests employed. We calculated FI and FQ, analyzing associations with trial characteristics through linear regression.

Results: We included 57 trials with a median sample size of 211 participants and a 12% median lost to follow-up. Most studies investigated anticholinergics (37/57, 65%). The median FI/FQ was 5/3.5%. Larger trials were less fragile (median FI 8; FQ 1.0%) compared to medium (FI: 4; FQ 2.5%) and small trials (FI: 4; FQ 8.3%). Double-blinded studies exhibited higher FQs (median 2.9%) than unblinded trials (6.7%). Primary and secondary outcomes had higher FIs (median 5 and 6, respectively) than adverse events (FI: 4). Each increase in 10 participants was associated with a +0.19 increase in FI ($p < 0.001$).

Conclusions: A change in outcome for a median of five participants, or 3.5% of the total sample size, could reverse the direction of statistical significance in OAB trials. Studies with larger sample sizes and efficacy outcomes from blinded trials were less fragile.

KEYWORDS

anticholinergics, B3 agonizts, clinical trials, fragility, fragility index, fragility quotient, mirabegron, overactive bladder, oxybutynin, systematic review

Abbreviations: CI, confidence interval; FI, fragility index; FQ, fragility quotient; IQR, interquartile range; OAB, overactive bladder; RCTs, randomized controlled trials.

1 | INTRODUCTION

Overactive bladder (OAB) syndrome, defined as urinary urgency with or without urge urinary incontinence and usually with frequency and nocturia, significantly disrupts patients' quality of life.[1,2] This condition is particularly burdensome for elderly patients, where it is associated with an increased risk of falls, fractures, and social isolation.[3,4] The cornerstone of management for OAB includes behavioral modifications, lifestyle adjustments, and pharmacological interventions.[5] Among pharmacotherapies, anticholinergics (or antimuscarinics) and beta-3 agonizts are frequently prescribed. While these medication classes have demonstrated efficacy in symptom management, some carry risks of adverse events, particularly cognitive effects in the elderly and cardiovascular issues in vulnerable populations.[6,7] In clinical trials, particularly those involving OAB medications, the robustness and reliability of study results are paramount. To quantify the strength of trial outcomes, Walsh et al. introduced the fragility index (FI) in 2014.[8] This index is computed as the smallest number of patients whose outcome status needs to be reversed to alter the statistical significance of a result. A smaller FI indicates a more fragile outcome, while the fragility quotient (FQ), calculated as FI divided by the total sample size and multiplied by 100, reflects this fragility in relation to the study's size. These metrics were a response to the recognition that statistical significance, often based on p-values, may not fully convey the strength of evidence, especially in trials with marginal results.[9] Thus the FI and FQ allow researchers to assess the vulnerability of a trial's findings to changes in patient outcomes, which is particularly salient in studies with small sample sizes or events.

The literature on fragility in urology clinical trials has consistently demonstrated low FI for various conditions, but assessments of fragility in OAB medication trials remain unexplored.[10–13] Furthermore, prior fragility studies have limited their scope to significant outcomes assessed by Fisher's exact test when calculating FI. Considering adverse effects associated with OAB medications, our analysis included both efficacy and adverse outcomes, applying fragility calculations to outcomes, significant and nonsignificant, using Fisher's exact tests, chi-squared tests, risk differences and odds ratios. Finally, we investigated what trial characteristics are associated with fragility. We hypothesized that trials with more rigorous designs, including double-blinding, larger sample sizes, and a focus on efficacy outcomes, would exhibit less fragility.

2 | METHODS

2.1 | Identification of trials

We identified trials published between January 2000 and August 2023 through a systematic search of PubMed, Web of Science, and Embase databases (Supporting Information S1: Table 1). Following a standardized protocol, we independently screened all identified studies. Eligible trials included randomized controlled trials (RCTs) with two parallel arms, reporting at least one binary outcome related to OAB medication efficacy or safety. We excluded crossover and cluster RCTs, as well as trials that analyzed previously published RCTs. Trials focusing exclusively on nonpharmacological interventions for OAB were also omitted.

2.2 | Data collection

Following a standardized data extraction protocol, five authors (NV, AMF, NH, UG, SP) independently extracted relevant data, which included trial identification details, publication year, publishing journal, journal impact factor, medication class, comparator type (placebo or active), total sample size (including in each arm), number lost to follow-up, study phase based on trial registration, and blinding status (single, double, or none). We recorded binary outcomes, including type, statistical test, and p value, and derived event numbers from reported proportions when not directly stated. Discrepancies in data extraction were adjudicated by an independent reviewer (KDL).

2.3 | Statistical analyses

The FI for each trial was calculated as described by Walsh et al.[8] To account for the size of the trials, we also calculated the FQ, which is the FI divided by the total sample size, multiplied by 100. Fragility calculations were computed using the "fragility" package in R.

Associations between FI and FQ with trial characteristics were evaluated using unadjusted linear regression. The distribution of FI/FQ values among categorical variables was assessed with the Kruskal-Wallis test, and for any significant findings, Dunn's test with Holm's adjustment for multiple comparisons to maintain the family-wise error rate.[14] Continuous variables were compared using the Wilcoxon rank sum test. A p-value threshold of <0.05 was predefined for statistical significance, and all tests were conducted as two-tailed. R

software (version 4.3.1) was used for all statistical analyses.

3 | RESULTS

3.1 | Trial and outcome characteristics

We included 57 trials reporting a total of 227 binary outcomes in our analysis (Figure 1). The median sample size was 211 participants (interquartile range [IQR]: 78–617; Table 1). The median loss to follow-up was 12% (IQR: 2%–18%). Most studies investigated anticholinergic medications (37/57, 65%), 14% (8/57) examined B3-agonizts, and 21% (12/57) studied medications other than anticholinergic or B3-agonizts.

Placebo-controlled trials comprised 65% (37/57) of the studies. Regarding study phase, 47% (27/57) were phase 3, 33% (19/57) were phase 4, and for 19% (11/57) the registered phase could not be located. Double-blinding was used in 81% (46/57) of studies; 8.8% (5/57) were unblinded and 5.3% (3/57) were single-blinded or did not report blinding status.

The median impact factor of publishing journals was 2.70 (IQR: 2.06–6.30). The *Journal of Urology* accounted for 23% (13/57) of publications, followed by *BJU International* (6/57, 11%) and *Urology* (4/57, 7%).

Each study reported a median of 2 binary outcomes (IQR: 1–6). Adverse events were the most reported outcome type (128/227, 57%), followed by primary (54/227, 24%) and secondary outcomes (45/227, 20%). Fisher's exact test (107/227, 47%) and chi-squared tests (94/227, 41%) were the most used statistical methods. Nonsignificant outcomes comprised 69% (156/227) of all outcomes.

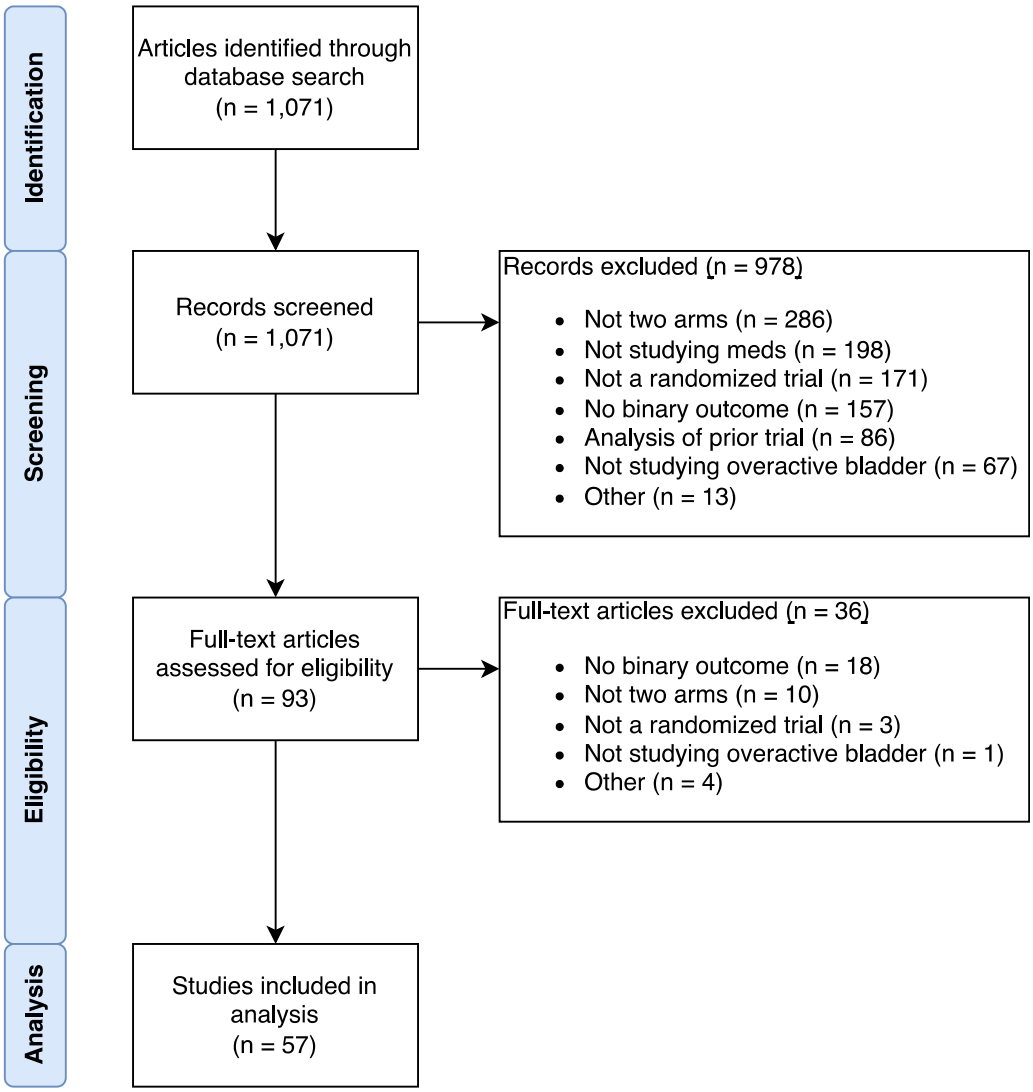

FIGURE 1 Flowchart of studies included per preferred reporting items for systematic reviews and meta-analysis.

TABLE 1 Characteristics of included studies ($N = 57$) and outcomes ($N = 225$).

Characteristic	N (%)
Sample size[a]	211 (78–617)
Percent lost to follow-up[a]	12 (2–18)
Lost to follow-up not reported[a]	6 (11)
Medication class	
Anticholinergic	37 (65)
Other	12 (21)
B3-agonist	8 (14)
Control arm presence	
Placebo controlled	37 (65)
Active comparator	20 (35)
Study phase	
Phase 3	27 (47)
Phase 4	19 (33)
Unable to locate	11 (19)
Blinding	
Double	46 (81)
Unblinded	5 (8.8)
Single	3 (5.3)
Not reported	3 (5.3)
Journal impact factor[a]	2.70 (2.06–6.30)
Top 10 journals	
Journal of Urology	13 (23)
BJU International	6 (11)
Urology	4 (7.0)
International Journal of Clinical Practice	3 (5.3)
Journal of the American Geriatrics Society	3 (5.3)
LUTS: Lower Urinary Tract Symptoms	3 (5.3)
Neurourology and Urodynamics	3 (5.3)
Clinical Drug Investigation	2 (3.5)
European Urology	2 (3.5)
Obstetrics & Gynecology	2 (3.5)
Binary outcomes per Study[a]	2 (1–6)
Fragility Index[a]	5 (3–7)
Fragility Quotient (%)[a]	3.5 (1.3–6.7)
Outcome type	
Adverse event	128 (56)
Primary	54 (24)
Secondary	45 (20)
Outcome test	
Fisher	107 (47)
Chi-squared	94 (41)
Risk difference	14 (6.2)
Odds ratio	12 (5.3)
Outcome significance	
Nonsignificant	156 (69)
Significant	71 (31)

[a]Median (IQR).

3.2 | Fragility overview

The median FI across all outcomes was 5 (IQR: 3–7; Figure 2), indicating that a median of five changes in outcome would alter the direction of statistical significance. The median FQ was 3.5% (IQR: 1.3%–6.7%), reflecting the proportionate fragility relative to trial sample size.

3.3 | Outcomes and trial size

Fragility across different study subgroups statistically differed by outcome types, sample sizes, and statistical significance for FIs, and by outcome type, sample size, medication class, trial blinding, and trial phase for FQs (Table 2).

Primary and secondary outcomes demonstrated higher FIs (medians, 5 and 6, respectively) and FQs (5.6% and 2.5%, respectively) than adverse event outcomes (FI median, 4; FQ median, 2.9%). Adverse events were more fragile than primary (FI $p = 0.020$; FQ $p = 0.003$) and secondary outcomes (FI $p = 0.004$), with significant pairwise comparisons depicted in Figure 3 FI and 4 FQ. Significant outcomes had a higher median FI of 6 compared to nonsignificant outcomes (FI: 5), with a corresponding FQ of 2.7% versus 3.6%, respectively, indicating significantly more fragility in nonsignificant outcomes (FI, $p = 0.014$).

Large trials were less fragile (FI: 8; FQ: 1.0%) compared to medium (FI: 4; FQ: 2.5%) and small trials (FI: 4; FQ: 8.3%). There were significant differences between large versus medium (FI, $p < 0.001$; FQ $p > 0.05$) and small trials (FI, $p < 0.001$). Journal impact factor was not significantly associated with fragility.

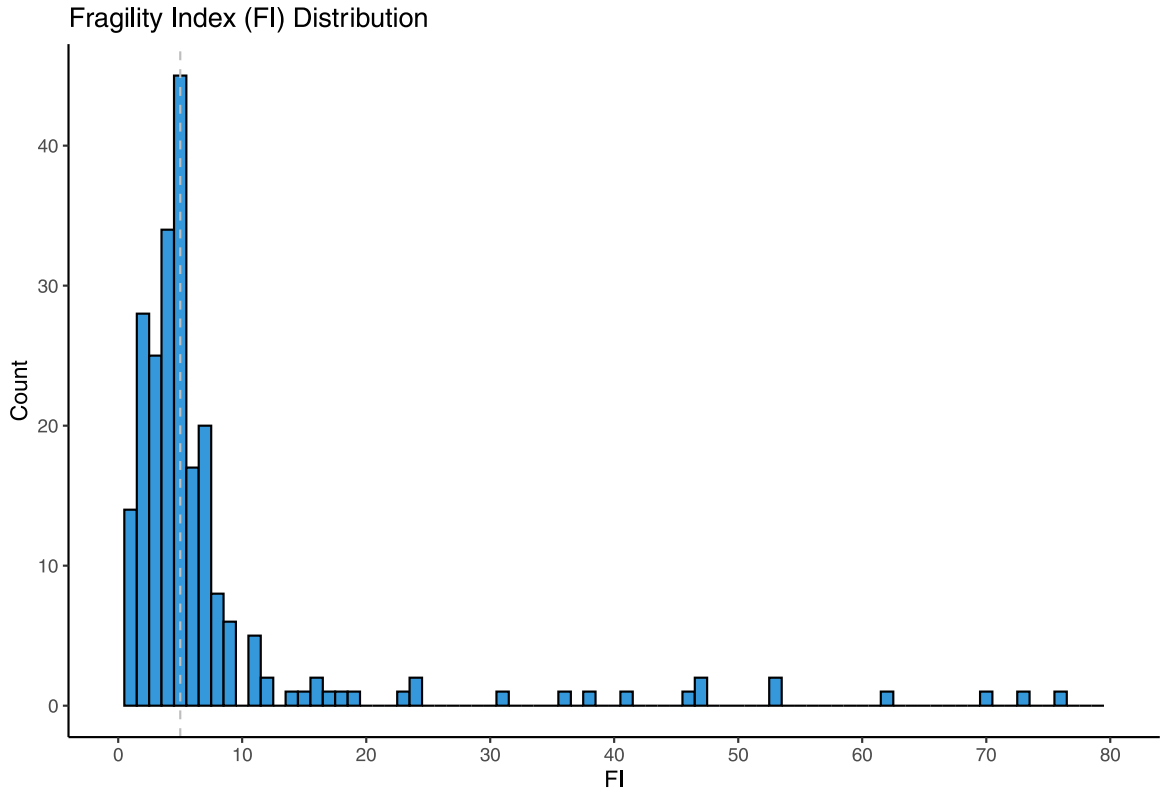

FIGURE 2 Distribution fragility indices across study outcomes.

3.4 | Medication class and trial design

Anticholinergics and B3 agonizts had similar FIs (both 5), but FQ was higher for B3 agonizts (4.5%) compared to anticholinergics (FQ 2.9%). Other medications had a lower FI (3) and higher FQ (5.6%). However, there were no significant differences after adjusting for multiple testing.

Unblinded trials had a median FI of 5, similar to double blinded trials. However, unblinded trials had a significantly higher FQ (6.7%) than that of double blinded trials (FQ 2.9%, $p < 0.001$). Comparison type (placebo vs. active comparator) was not significantly associated with fragility.

Phase III trials reported a median FI of 5 and FQ of 2.8%, while phase IV trials showed a lower FI (4) and higher FQ (4.8%). Trials with unspecified phases had a median FI of 5 and FQ of 4.6%. Phase III trials were significantly less fragile than phase IV (FQ, $p = 0.006$) and unspecified phase trials (FQ, $p < 0.001$).

3.5 | Linear associations

We employed linear regression to assess associations between study/outcome characteristics and fragility (Table 3). For each increase of 10 patients in sample size, there was an associated +0.19 increase in FI (95% Confidence Interval [CI] 0.15, 0.24; $p < 0.001$) and a −0.06% decrease in FQ (95% CI: −0.08, −0.05; $p < 0.001$). Each additional binary outcome reported was associated with a −0.45 reduction in FI (95% CI: −0.73, −0.17; $p = 0.002$) and a + 0.18% elevation in FQ (95% CI 0.09, 0.27; $p < 0.001$). Relative to adverse events, primary outcomes were linked to a + 5.8 greater FI (95% CI: 2.1, 9.5; $p = 0.002$), and secondary outcomes to a + 6.5 increase (95% CI: 2.6, 10; $p = 0.001$). Statistically significant results were associated with a + 9.7 rise in FI (95% CI: 6.5, 13; $p < 0.001$), but not significantly with FQ. Studies with placebo controls, versus active comparators, had a +5.2 higher FI (95% CI 2.0, 8.3; $p = 0.001$) with no notable change in FQ. Compared to double-blinded studies, unblinded studies were associated with a + 2.8% increase in FQ (95% CI: 1.3, 4.2; $p < 0.001$) and single-blinded studies exhibited a +2.7% increase (95% CI: 0.24, 5.2; $p = 0.032$).

In a focused sub-analysis of efficacy-related outcomes (primary and secondary; Supporting Information S1: Table 2), a unit increase in Journal Impact Factor was tied to a +0.43% rise in FQ (95% CI 0.18, 0.68; $p < 0.001$), while statistically significant outcomes were associated with a −2.1% decrease in FQ (95% CI: −3.6, −0.59; $p = 0.007$). Compared to Phase III trials, Phase IV trials were associated with an −8.3 decrease in FI (95% CI: −16, −0.36; $p = 0.041$), and trials without registration reported with a −13 reduction (95% CI: −21, −4.5; $p = 0.002$). Unblinded trials had a −9.8 lower FI (95% CI: −18, −1.1; $p = 0.027$). The associations

TABLE 2 Fragility index and fragility quotient by different study subgroups.

Characteristic	Fragility index		Fragility quotient (%)	
	Median (IQR)	*p* value[a]	Median (IQR)	*p* value[a]
Outcome type		0.001		0.002
Adverse event	4 (3–6)		2.9 (1.3–5.6)	
Primary	5 (4–7)		5.6 (2.1–8.5)	
Secondary	6 (4–11)		2.5 (1.0–4.9)	
Sample size		<0.001		<0.001
Large (>=500)	8 (5–31)		1.0 (0.6–3.7)	
Medium (100–499)	4 (3–6)		2.5 (1.2–3.7)	
Small (<100)	4 (3–5)		8.3 (5.6–10.6)	
Journal impact		0.3		0.5
High (>=3)	5 (3–7)		2.8 (1.0–7.8)	
Low (<3)	5 (3–7)		3.6 (1.9–6.2)	
Comparison group		0.081		0.093
Active comparator	5 (3–6)		3.6 (2.1–6.6)	
Placebo controlled	5 (3–8)		2.8 (0.8–7.4)	
Medication class		0.11		0.036
Anticholinergic	5 (3–7)		2.9 (0.9–6.8)	
B3 Agonist	5 (3–6)		4.5 (2.2–5.5)	
Other	3 (2–6)		5.6 (2.0–7.4)	
Trial blinding		0.2		<0.001
Double	5 (3–7)		2.9 (1.0–6.6)	
Not reported	5 (4–7)		2.1 (2.0–2.8)	
Single	3 (3–5)		6.3 (5.0–8.3)	
Unblinded	5 (2–7)		6.7 (4.8–8.7)	
Significance		0.014		0.086
Nonsignificant	5 (3–6)		3.6 (1.5–7.3)	
Significant	6 (3–17)		2.7 (1.0–5.6)	
Trial phase		0.2		<0.001
III	5 (3–8)		2.8 (0.9–4.5)	
IV	4 (3–7)		4.8 (1.3–7.1)	
Unable to locate	5 (4–6)		4.6 (2.1–10.6)	

[a]Kruskal-Wallis rank sum test; Wilcoxon rank sum test.

with other study characteristics were consistent with the overall findings.

4 | DISCUSSION

We systematically reviewed and analyzed the fragility of clinical trial outcomes for OAB medications. Overall, clinical trials of OAB medications were fragile, with a median change of 5 participants or 3.5% of the study population reversing the direction of statistical significance, indicating that their outcomes are vulnerable to small fluctuations in trial events. Larger sample sizes and primary outcome reporting were associated with less fragile outcomes. Conversely, smaller trials, those reporting adverse events, and studies with less stringent blinding procedures were more fragile. Our linear regression analyses underscored that larger trials and

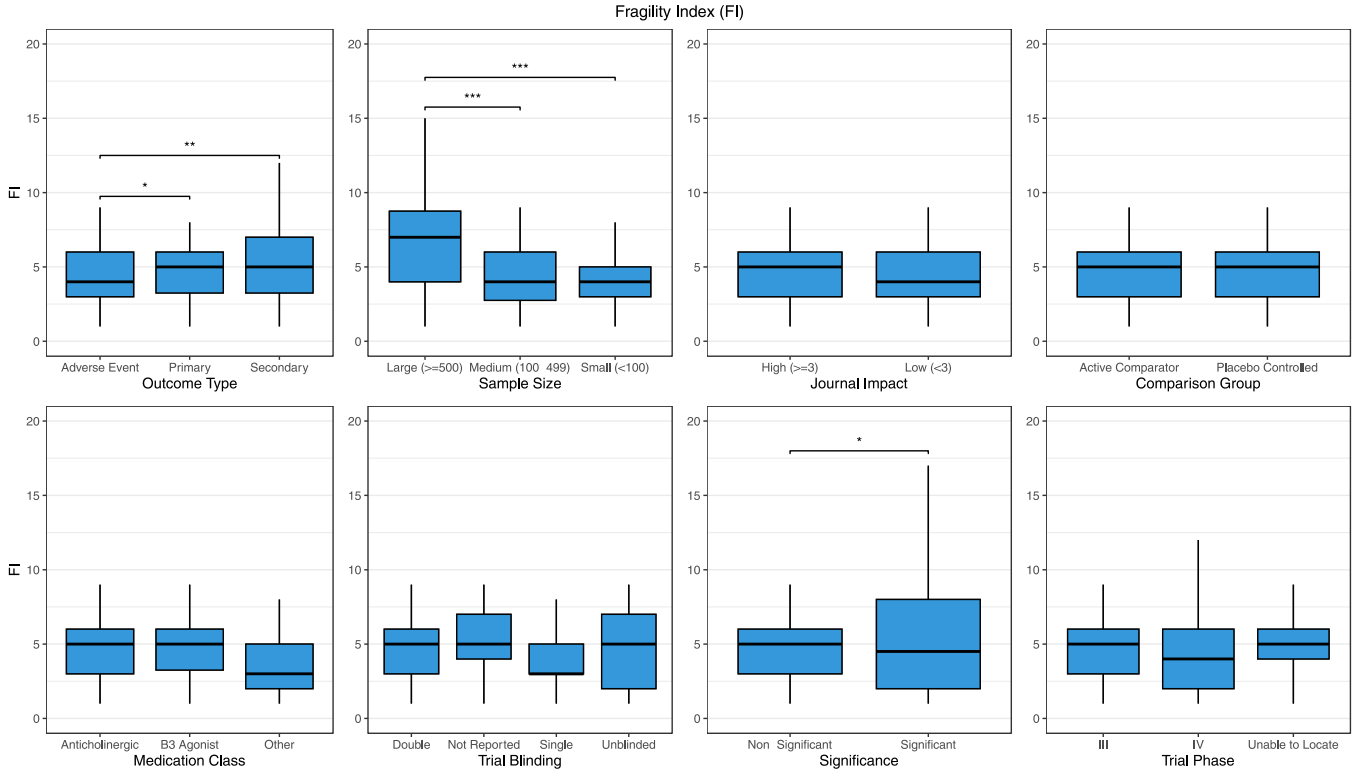

FIGURE 3 Differences in fragility index by study characteristics. Brackets denote statistically significant post-hoc comparisons after adjustment with Holm's method.

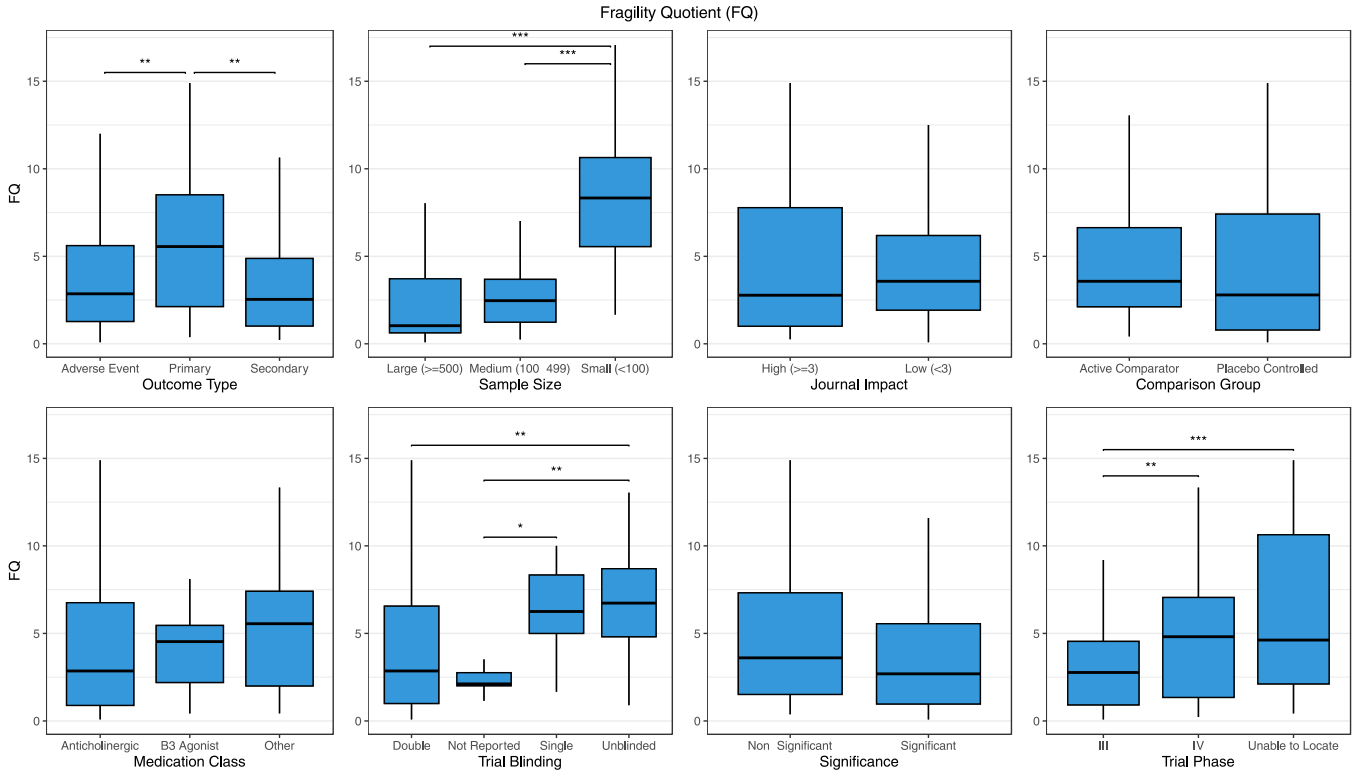

FIGURE 4 Differences in fragility quotient by study characteristics. Brackets denote statistically significant post-hoc comparisons after adjustment with Holm's method.

TABLE 3 Association between study characteristics and the fragility using linear regression.

Characteristic	Fragility index		Fragility quotient (%)	
	Beta[a]	*p* value	Beta[a]	*p* value
Sample size + 10	0.19 (0.15, 0.24)	<0.001	−0.06 (−0.08, −0.05)	<0.001
Number of binary outcomes	−0.45 (−0.73, −0.17)	0.002	0.18 (0.09, 0.27)	<0.001
Journal impact factor	0.37 (−0.16, 0.90)	0.2	−0.06 (−0.23, 0.11)	0.5
Percent lost to follow-up	−0.24 (−0.39, −0.09)	0.001	−0.05 (−0.10, 0.01)	0.075
Statistically significant outcome	9.7 (6.5, 13)	<0.001	−0.85 (−1.9, 0.23)	0.12
Outcome type				
Adverse event	1.00		1.00	
Primary	5.8 (2.1, 9.5)	0.002	2.0 (0.78, 3.2)	0.001
Secondary	6.5 (2.6, 10)	0.001	−0.57 (−1.8, 0.70)	0.4
Medication class				
Anticholinergic	1.00		1.00	
B3 Agonist	−4.4 (−8.1, −0.74)	0.019	0.60 (−0.59, 1.8)	0.3
Other	−2.7 (−8.1, 2.7)	0.3	1.1 (−0.63, 2.9)	0.2
Trial phase				
III	1.00		1.00	
IV	−2.3 (−5.8, 1.3)	0.2	1.8 (0.75, 2.9)	0.001
Unable to locate	−2.6 (−6.7, 1.5)	0.2	3.2 (1.9, 4.4)	<0.001
Trial blinding				
Double	1.00		1.00	
Not reported	−3.6 (−9.3, 2.0)	0.2	−1.8 (−3.5, −0.01)	0.048
Single	−5.2 (−13, 2.8)	0.2	2.7 (0.24, 5.2)	0.032
Unblinded	−4.7 (−9.4, 0.02)	0.051	2.8 (1.3, 4.2)	<0.001
Comparison group				
Active comparator	1.00		1.00	
Placebo controlled	5.2 (2.0, 8.3)	0.001	−0.18 (−1.2, 0.85)	0.7

[a]Expected change in Fragility per unit increase in covariate.

efficacy outcomes from blinded studies were less fragile. Our study is the first fragility analysis of OAB medications, including both significant and nonsignificant binary outcomes. Given our findings, we advocate for inclusion of FI/FQ in the reporting of future clinical trials.

Fragility analyses in urology have reported median FIs as low as 2 for low-intensity extracorporeal shockwave therapy in erectile dysfunction, 3 in urologic oncology,[15] 3.5 for medical expulsive treatment of ureteral stones,[16] 4 in bladder and bowel dysfunction literature,[17] and 5 in pediatric urology meta-analyses.[18] Trials across other medical specialties have been noted to have similarly fragile outcomes. Shoulder arthroplasty and oncology trials have been shown to have a median FI of 6 and 2, respectively.[19,20] Similar fragility is seen in orthopedic oncology and bariatric surgery trials, each with a median FI of 2,[21,22] while critical care trials reveal an identical trend.[23] In contrast, cardiac disease and heart failure trials often display greater FI; for instance, the median FI for trials supporting acute coronary syndrome guidelines is 12.[24,25]

In a meta-analysis spanning multiple research fields, Holek et al. (2020) identified a high level of fragility in clinical trials, noting that factors such as higher journal impact factor, larger sample sizes, and greater effect sizes were associated with an increased FI.[26] Our results corroborate these findings, revealing that these same factors, alongside more outcome events and a lower p-value, contribute to a higher FI in OAB trials. A

particularly concerning point raised by Holek et al. is the proportion of trials where the number of patients lost to follow-up exceeds the FI, which was the case in 44% of trials in their meta-analysis, suggesting that the significance of the trial results could be reversed if outcomes of patients lost to follow-up differed from those analyzed. Our study observed this in 26.5% of studies (13/49 which reported follow-up losses), indicating a lower but still significant impact of lost follow-up on interpretations of trial results.

The FI, while providing an easily interpretable measure of trial robustness, has been subject to criticism. The FI's focus on the statistical significance of results may detract from the broader clinical context and the importance of understanding data holistically.[27] Furthermore, a trial's FI may reflect meticulous design aimed at efficiency rather than indicate a trial's susceptibility to outcome changes.[28,29] Concerns also arise from the FI's potential to misguide interpretations, ignoring the ethical imperative of minimizing participant exposure once clinical equipoise is lost.[30] Therefore, while FI can highlight the need for cautious interpretation of trial outcomes, it should not be isolated from the comprehensive statistical and clinical appraisal necessary for informed decision-making in healthcare.[27,30] Some authors have instead recommended the adoption of modern approaches, such as Bayesian and Monte Carlo sensitivity analyses, which provide more sophisticated assessments of robustness within clinical trials.[27]

Our study, while offering insights into OAB trial robustness, has important limitations. Primarily, FI calculations were confined to binary outcomes, leading to the exclusion of a substantial number of studies ($n = 175$) that reported continuous outcomes. This restriction may result in an incomplete representation of fragility across the full spectrum of OAB trials. Additionally, our methodology did not extend to time-to-event outcomes due to the inapplicability of fragility calculations for such data, which omits a critical aspect of many clinical trials that could influence the interpretation of their results. Strengths of our study include a comprehensive systematic review encompassing a broad range of test types and the inclusion of nonsignificant and secondary/adverse event outcomes, which provides a holistic view of fragility. Additionally, application of the FQ, which is the FI normalized for sample size, allowed interpretation of fragility relative to trial size.

We advocate for consideration of not only the statistical significance of reported results but also their effect sizes and confidence intervals.[31,32] These considerations can provide a more nuanced understanding of study outcomes, beyond a p-value threshold. From the perspective of clinical trial reporting, our review highlights deficiencies that should be addressed to improve transparency in published research. Notably, 11% of trials we reviewed did not report losses to follow-up, 19% lacked trial registration details, and 5.3% did not report blinding status. This highlights the need for more rigorous reporting standard in clinical trials, thereby enabling a thorough evaluation of reported outcomes and reinforcing their credibility.

5 | CONCLUSIONS

Our systematic review revealed that a median shift in outcome for five individuals, or 3.5% of the total sample size, would be sufficient to reverse the direction of statistical significance in OAB medication trials. We observed that larger trials and those with primary outcomes tended to be less fragile, while smaller studies, adverse event reports, and less stringent blinding procedures in efficacy-focused outcomes were associated with greater fragility. Nuanced interpretation of trial results, including effect sizes and confidence intervals, and comprehensive reporting of trial details are next steps to strengthen research robustness and transparency.

AUTHOR CONTRIBUTIONS

Kevin D. Li: Data collection, analysis, manuscript writing. **Benjamin N. Breyer**, **Hiren V. Patel**: Supervision, concept, critical revisions. **Nikit Venishetty**, **Adrian M. Fernandez**, **Nizar Hakam**, **Umar Ghaffar**, **Shiv Gupta**: Data collection, manuscript revision.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. All authors have declared that no financial support was received.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest relevant to the content of this review.

DATA AVAILABILITY STATEMENT

Data and analyses for this study are available in a public repository (github.com/kevndli/oab-fragility).

ETHICS STATEMENT

Ethical approval was not required for this study.

ORCID

Kevin D. Li http://orcid.org/0000-0003-0550-0250
Adrian M. Fernandez http://orcid.org/0000-0002-9918-5783

REFERENCES

1. Stewart W, Van Rooyen J, Cundiff G, et al. Prevalence and burden of overactive bladder in the United States. *World J Urol*. 2003;20(6):327-336. doi:10.1007/s00345-002-0301-4
2. Coyne KS, Wein AJ, Tubaro A, et al. The burden of lower urinary tract symptoms: evaluating the effect of LUTS on health-related quality of life, anxiety and depression: EpiLUTS. *BJU Int*. 2009;103(s3):4-11. doi:10.1111/j.1464-410X.2009.08371.x
3. Ganz ML, Liu J, Zou KH, Bhagnani T, Luo X. Real-world characteristics of elderly patients with overactive bladder in the United States. *Curr Med Res Opin*. 2016;32(12):1997-2005. doi:10.1080/03007995.2016.1226167
4. Soliman Y, Meyer R, Baum N. Falls in the elderly secondary to urinary symptoms. *Rev Urol*. 2016;18(1):28-32.
5. Lightner DJ, Gomelsky A, Souter L, Vasavada SP. Diagnosis and treatment of overactive bladder (non-neurogenic) in adults: AUA/SUFU guideline amendment 2019. *J Urol*. 2019;202(3):558-563. doi:10.1097/JU.0000000000000309
6. Wagg A, Staskin D, Engel E, Herschorn S, Kristy RM, Schermer CR. Efficacy, safety, and tolerability of mirabegron in patients aged ≥65yr with overactive bladder wet: a phase IV, double-blind, randomised, placebo-controlled study (PILLAR). *Eur Urol*. 2020;77(2):211-220. doi:10.1016/j.eururo.2019.10.002
7. Farag F, Sakalis VI, Arteaga SM, et al. What are the short-term benefits and potential harms of therapeutic modalities for the management of overactive bladder syndrome in women? A review of evidence under the auspices of the european association of urology, female non-neurogenic lower urinary tract symptoms guidelines panel. *Eur Urol*. 2023;84(3):302-312. doi:10.1016/j.eururo.2023.05.014
8. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. *J Clin Epidemiol*. 2014;67(6):622-628. doi:10.1016/j.jclinepi.2013.10.019
9. Scientists rise up against statistical significance. Accessed December 28 2023. https://www-nature-com.ucsf.idm.oclc.org/articles/d41586-019-00857-9
10. Rickard M, Lorenzo AJ, Hannick JH, Blais AS, Koyle MA, Bägli DJ. Over-reliance on P values in urology: fragility of findings in the hydronephrosis literature calls for systematic reporting of robustness indicators. *Urology*. 2019;133:204-210. doi:10.1016/j.urology.2019.03.045
11. Gnech M, Lovatt CA, McGrath M, et al. Quality of reporting and fragility index for randomized controlled trials in the vesicoureteral reflux literature: where do we stand. *J Pediatr Urol*. 2019;15(3):204-212. doi:10.1016/j.jpurol.2019.02.014
12. Narayan VM, Gandhi S, Chrouser K, Evaniew N, Dahm P. The fragility of statistically significant findings from randomised controlled trials in the urological literature. *BJU Int*. 2018;122(1):160-166. doi:10.1111/bju.14210
13. Li B, Kong I, McGrath M, Farrokhyar F, Braga LH. Evaluating the literature on preoperative androgen stimulation for hypospadias repair using the fragility index – can we trust observational studies. *J Pediatr Urol*. 2021;17(5):661-669. doi:10.1016/j.jpurol.2021.07.027
14. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *Am J Public Health*. 1996;86(5):726-728.
15. Chertin L, Burman I, Haifler M. Urologic oncology randomized controlled trials are frequently fragile - a review of the urology literature. *Urol Oncol*. 2021;39(10):735. doi:10.1016/j.urolonc.2021.04.037
16. Tzelves L, Chatzikrachtis N, Lazarou L, et al. Fragility index of urological literature regarding medical expulsive treatment. *World J Urol*. 2021;39(10):3741-3746. doi:10.1007/s00345-021-03725-2
17. Rickard M, Keefe DT, Drysdale E, et al. Trends and relevance in the bladder and bowel dysfunction literature: PlumX metrics contrasted with fragility indicators. *J Pediatr Urol*. 2020;16(4):477.e1-477.e7. doi:10.1016/j.jpurol.2020.06.015
18. Anand S, Kainth D. Fragility index of recently published meta-analyses in pediatric urology: A striking observation. *Cureus*. 2021;13(7):e16225. doi:10.7759/cureus.16225
19. Mccormick KL, Tedesco LJ, Swindell HW, Forrester LA, Jobin CM, Levine WN. Statistical fragility of randomized clinical trials in shoulder arthroplasty. *J Shoulder Elbow Surg*. 2021;30(8):1787-1793. doi:10.1016/j.jse.2020.10.028
20. Del Paggio JC, Tannock IF. The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a retrospective analysis. *Lancet Oncol*. 2019;20(8):1065-1069. doi:10.1016/S1470-2045(19)30338-9
21. Forrester LA, Jang E, Lawson MM AC, Capi A, Tyler WK. Statistical fragility of surgical and procedural clinical trials in orthopaedic oncology. *J Am Acad Orthop Surg Global Res Rev*. 2020;4(6):e19.00152. doi:10.5435/JAAOSGlobal-D-19-00152
22. Lee Y, Samarasinghe Y, Chen LH, et al. Fragility of statistically significant outcomes in randomized trials comparing bariatric surgeries. *Int J Obes*. 2023;47(7):546-553. doi:10.1038/s41366-023-01298-1
23. Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The fragility index in multicenter randomized controlled critical care trials. *Crit Care Med*. 2016;44(7):1278-1284. doi:10.1097/CCM.0000000000001670
24. Gonzalez-Del-Hoyo M, Mas-Llado C, Blaya-Peña L, Siquier-Padilla J, Peral V, Rossello X. The fragility index in randomized clinical trials supporting clinical practice guidelines for acute coronary syndrome: measuring robustness from a different perspective. *Eur Heart J Acute Cardiovasc Care*. 2023;12(6):386-390. doi:10.1093/ehjacc/zuad021
25. Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg*. 2019;154(1):74-79. doi:10.1001/jamasurg.2018.4318
26. Holek M, Bdair F, Khan M, et al. Fragility of clinical trials across research fields: a synthesis of methodological reviews. *Contemp Clin Trials*. 2020;97:106151. doi:10.1016/j.cct.2020.106151
27. Potter GE. Dismantling the fragility index: a demonstration of statistical reasoning. *Stat Med*. 2020;39(26):3720-3731. doi:10.1002/sim.8689
28. Carter RE, McKie PM, Storlie CB. The fragility index: a p-value in sheep's clothing? *Eur Heart J*. 2017;38(5):346-348. doi:10.1093/eurheartj/ehw495
29. Machado T, Duarte GS, Gonçalves N, Ferreira JJ, Costa J. A critique of the fragility index. *Lancet Oncol*. 2019;20(10):e553. doi:10.1016/S1470-2045(19)30581-9

30. Stensland KD, Daignault-Newton S, Skolarus TA. Designing lean, efficient clinical trials is an ethical imperative: the fragility index should not be used in the design of randomized clinical trials. *Urol Oncol*. 2021;39(10):738-739. doi:10.1016/j.urolonc.2021.07.004
31. Muehlemann N, Zhou T, Mukherjee R, Hossain MI, Roychoudhury S, Russek-Cohen E. A tutorial on modern Bayesian methods in clinical trials. *Ther Innov Regul Sci*. 2023;57(3):402-416. doi:10.1007/s43441-023-00515-3
32. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean J Anesthesiol*. 2016;69(6):555-562. doi:10.4097/kjae.2016.69.6.555

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li KD, Venishetty N, Fernandez AM, et al. Fragility of overactive bladder medication clinical trials: a systematic review. *Neurourol Urodyn*. 2024;43:1523-1533. doi:10.1002/nau.25468