Spine

SPINE Volume 46, Number 3, pp 198–208

LITERATURE REVIEW

Fragility Analysis of Statistically Significant Outcomes of Randomized Control Trials in Spine Surgery

A Systematic Review

Sathish Muthu, MS Ortho[a,1] and Eswar Ramakrishnan, MBBS[b,1]

Study Design. Systematic review.

Objectives. The aim of this study was to assess the robustness of statistically significant outcomes from randomized control trials (RCTs) in spine surgery using Fragility Index (FI) which is a novel metric measuring the number of events upon which statistical significance of the outcome depends.

Summary of Background Data. Many trials in Spine surgery were characterized by fewer outcome events along with small sample size. FI helps us identify the robustness of the results from such studies with statistically significant dichotomous outcomes.

Methods. We conducted independent and in duplicate, a systematic review of published RCTs in spine surgery from PubMed Central, Embase, and Cochrane Database. RCTs with 1:1 prospective study design and reporting statistically significant dichotomous primary or secondary outcomes were included. FI was calculated for each RCT and its correlation with various factors was analyzed.

Results. Seventy trials met inclusion criteria with a median sample size of 133 (interquartile range [IQR]: 80–218) and median reported events per trial was 38 (IQR: 13–94). The median FI score was 2 (IQR: 0–5), which means if we switch two patients from nonevent to event, the statistical significance of the outcome is lost. The FI score was less than the number of patients lost to follow-up in 28 of 70 trials. The FI score was found to positively correlated with sample size ($r = 0.431$, $P = 0.001$), total number of outcome events ($r = 0.305$, $P = 0.01$) while negatively correlated with P value ($r = -0.392$, $P = 0.001$). Funding, journal impact-factor, risk of bias domains, and year of publication did not have a significant correlation.

Conclusion. Statistically significant dichotomous outcomes reported in spine surgery RCTs are more often fragile and outcomes of the patients lost to follow-up could have changed the significance of results and hence it needs caution before transcending their results into clinical application. The addition of FI in routine reporting of RCTs would guide readers on the robustness of the statistical significance of outcomes. RCTs with FI ≥ 5 without any patient lost to follow-up can be considered to have clinically robust results.

Key words: evidence based medicine, Fragility Index, *P* value, randomized controlled trial, statistical data interpretation.

Level of Evidence: 1

Spine 2021;46:198–208

In the era of evidence-based medicine, randomized control trials (RCTs) form the keystone based on which clinical decisions are made and treatment protocols are formulated.[1] However, RCTs involving spine surgery often provide us with discordant results.[2–4] Although there are stringent protocols for the conduction and reporting of RCTs, little attention is given to evaluate the robustness of the significance of its outcome events.[5] Trials with reduced numbers of outcome events are at high risk of making large treatment effects, especially when sample sizes are also small.[6,7]

The Fragility Index (FI) is a novel metric developed to assess the robustness of the statistically significant dichotomous outcomes.[8] The FI is defined as the minimum number of patients who have to be changed from a nonevent to an event in the treatment group to lose the statistical significance of the dichotomous outcome analyzed such as fusion, adjacent segment disease, or pedicle breach.

For example, in an RCT by Glassman *et al*,[9] 102 patients undergoing lumbar spine fusion older than 60 years were randomized to receive either posterolateral lumbar fusion

From the [a]Government Hospital, Velayuthampalayam, Karur, Tamil Nadu, India; and [b]Institute of Orthopaedics and Traumatology, Madras Medical College & Rajiv Gandhi Government General Hospital, Chennai, Tamil Nadu, India.

[1]Researcher, Orthopaedic Research Group, Coimbatore, Tamil Nadu, India.

Acknowledgment date: March 29, 2020. First revision date: April 22, 2020. Acceptance date: June 5, 2020.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work.

No relevant financial activities outside the submitted work.

Address correspondence and reprint requests to Dr. Sathish Muthu, MS Ortho, Assistant Surgeon, Government Hospital, Velayuthampalayam, Karur 639117, Tamil Nadu, India; E-mail: drsathishmuthu@gmail.com

DOI: 10.1097/BRS.0000000000003645

TABLE 1. Fragility Index Calculation Example, Based on Trial by Glassman *et al*[9]

Sample Size (n = 102)	Complications	No Complications	*P*
Original results, no. (%)			
Intervention group	8 (16)	42 (84)	0.014
Control group	20 (38.5)	32 (61.5)	
First step of Fragility Index Calculation			
Intervention group	9	41	0.028
Control group	20	32	
Second step of Fragility Index Calculation			
Intervention group	10	40	0.051
Control group	20	32	

with rhBMP-2 in absorbable collagen sponge or iliac crest bone graft (ICBG). In this trial, eight complications were noted in the rhBMP-2 group, whereas the ICBG group had 20 complications. This difference was statistically significant ($P = 0.014$), but it would have been completely insignificant if just two more patients in the rhBMP-2 group had complications ($P = 0.051$) as shown in Table 1. Thus, the FI for this event outcome is 2, which means if two more events occurred in rhBMP-2 group the significance of the result is lost.

Many trials in Spine surgery were characterized by fewer outcome events along with a small sample size.[10] Hence the objective of this review is to assess the robustness of the statistically significant outcomes in RCTs of Spine surgery interventions by using FI and also analyze the factors associated with FI.

MATERIALS AND METHODS

Our methodology and reporting of the systematic review follows PRISMA[11] and AMSTAR 2[12] guidelines which consist of a 27-item checklist and 16-point assessment respectively to help authors improve the conduction and reporting of systematic reviews and meta-analyses.

Inclusion Criteria

To be included in our study, a study should meet the following criteria:

1. The study should be an RCT with 1:1 parallel two-arm design.
2. The study must be related to spine surgery involving preoperative or intraoperative or postoperative variables.
3. The study must have a dichotomous primary or secondary outcome.

Exclusion Criteria

1. Studies not involving human subjects.
2. Studies with continuous variable outcomes like pain scores, Oswestry Disability Index scores, time to union without predefined clinical success criteria.
3. Studies that did not report a statistically significant primary or secondary outcome measure.

Study Identification

We performed a computerized search of PubMed Central, Embase, and the Cochrane Database with the following terms and Boolean operators: (''spine'' OR ''spinal'') AND (''surgery'' OR ''methods'' OR ''procedure'' OR ''fracture'' OR ''infection'' OR ''deformity''). The results of the search were filtered based on the publication type to isolate RCTs. No language restriction was applied. The bibliography of each study was reviewed by both the authors to look for additional relevant studies. Both the authors independently reviewed the title of each article retrieved from the search for its relevance and excluded studies with identified reasons as mentioned in the flow diagram (Figure 1). After title screening, abstract and full-text screening was done by both the authors independently. Any discrepancy was settled by consensus. The agreement between two authors at each stage of screening was assessed by weighted kappa scores.[13] An interclass correlation coefficient was used for quality appraisal.

Assessment of Risk of Bias

Each eligible study was independently reviewed by both the authors for methodological quality with the Cochrane Collaboration Risk of Bias tool[14] which has eight domains of assessment as shown in Table 2.

Data Extraction

For every eligible study, the relevant data were extracted in duplicate with discrepancies resolved by consensus. We collected the statistically significant dichotomous outcome for every study included in the analysis. For studies reporting more than one dichotomous outcome, we chose the primary outcome of the study or the most critical outcome for decision making based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.[15] Extracted data include: journal name along with its impact factor from Thomson Reuters Journal Impact Factor, publication year, first author, funding

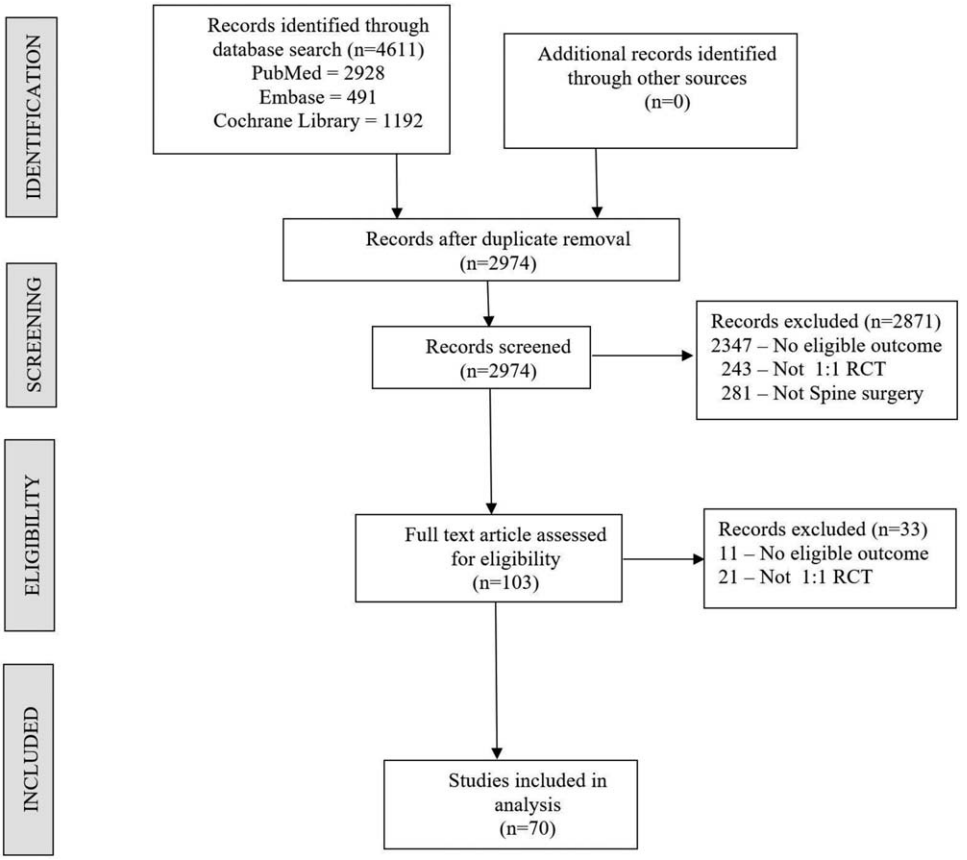

Figure 1. PRISMA flow diagram for the selection of studies into analysis and reason for exclusion of the articles.

source, sample size of the study with allocation for each arm and losses to final follow-up for each arm, number of events in each arm, reported *P* value, and the statistical test used.

Fragility Analysis

We calculated independent and in duplicate, the FI for each dichotomous outcome using an online FI calculation tool by Kane.[16] The least number of events to be added to the arm with the least reported events keeping the sample size constant, to obtain a *P* value $\geq$to 0.05 is the FI. Fragility Quotient (FQ) provides a method to evaluate the fragility relative to the sample size of the study. FQ is calculated by dividing the FI score by the total sample size of the study.

Descriptive statistics were determined using the Statistical Package for Social Sciences (SPSS®), Version 25 (SPSS Inc., IBM Corp., Armonk, NY). We identified the median FI among the identified studies and its correlation with variables like year of publication, sample size, events in the reported outcome, *P* value, funding, risk of bias domains, and journal impact factor. We evaluated correlation using the Pearson correlation coefficient. All tests of significance were two-tailed and a *P* value <0.05 was considered significant.

TABLE 2. Risk of Bias of the Included RCTs; n = 70 Trials

Characteristic	Category Risk Level Frequency (%)		
	Low Risk	High Risk	Unclear
Randomization	68 (97.1)	00 (0)	02 (2.8)
Allocation concealment	56 (80.0)	01 (1.4)	13 (18.5)
Patient masking	36 (51.4)	05 (7.1)	29 (41.4)
Surgeon masking	24 (34.2)	14 (20.0)	32 (45.7)
Outcome assessor masking	39 (55.7)	08 (11.4)	23 (32.8)
Incomplete outcome reporting	48 (68.5)	12 (17.1)	10 (14.2)
Selective reporting	51 (72.8)	02 (2.8)	17 (24.2)
Other source	60 (85.7)	03 (4.2)	07 (10.0)

RCT indicates randomized controlled trial.

Patients lost to follow-up remain a major confounding variable influencing the outcome in either arm of the trial. Hence, we also evaluated the patients lost to follow-up in the individual studies and compared it to their FI to further determine the fragility of their results. Patients lost to follow-up are defined as the difference between the number of patients randomized to the group and the number of patients evaluated for the reported outcome.

Source of Funding
No source of funding was utilized for this study.

RESULTS

Study Identification
Four thousand six hundred and eleven potentially relevant articles were identified: 2928 (63.5%) from PubMed Central, 491 (10.6%) from Embase, and 1192 (25.9%) from Cochrane Database from initial search out of which 1637 duplicates were removed and title and abstract screening were done for a total of 2974 articles and 103 articles were found eligible for full-text review and 70 articles published between 2002 and 2019 were included in the analysis as shown in Figure 1. Agreement between the authors for the title, abstract, and full-text screening for study identification were substantially high. (k = 0.84, 95% confidence interval [CI] = 0.79–0.88; k = 0.86, 95% CI = 0.81–0.90; k = 0.94, 95% CI = 0.87–0.96, respectively).

The most common reason for exclusion of RCTs from analysis being utilization of nondichotomous variables to report their outcomes like a time-to-event variable (time to fusion, time to complication) or ordinal variable (Visual Analog Scale score, ODI score) without predefined clinical success cutoff values in their trial protocol which prevented from considering them as a dichotomous outcome. Table 3 [9,17–85] lists the included RCTs, year and journal of publication, total sample size, total reported events in control and study group, *P* value of the reported dichotomous outcome and calculated FI and FQ.

Trial Characteristics and Outcomes
The characteristics of the included trials are shown in Table 4. The median sample size of the included trials was 133 (interquartile range [IQR]: 80–218) and 14.6% (n = 1933) of the patients were lost to follow-up across trials. The median journal impact factor was 2.79 (IQR 2.63–2.79). On considering the quality of the included trials there was a low risk of bias in sequence generation and allocation concealment in 68 (97.1%) and 56 (80%), respectively. Investigators blinded surgeons in 24 (34.2%), patients in 36 (51.4%), and outcome assessors in 39 (55.7%) as shown in Table 2.

Of the 70 outcomes analyzed, 44 (62.8%) were primary, whereas the remaining 26 (37.1%) were secondary outcomes. The median reported events per trial was 38 (IQR: 13–94). Thirty-seven (52.8%) of the included trials were funded for their research.

Fragility Analysis
Distribution of FI of the included trials and patients lost to follow-up in them are shown in Figure 2. The median FI of the 70 included trials was two events (IQR 0–5) which shows that by adding two events to one of the arms of the trial, the significance of the results obtained is lost. Nineteen outcomes (27.1%) lost their statistical significance once we recalculated their *P* values using a two-sided Fischer exact test. Hence, they had an FI of zero. In 37 of 70 trials, patients were lost to follow-up, of which in 28 trials, the number of patients lost to follow-up exceeded the FI. The median FQ score was 0.0148 (IQR 0–0.033)

FI had a significant inverse correlation with the reported *P* value of the outcomes ($r = -0.392$, $P = 0.0011$). Increasing FI values were significantly correlated with smaller reported *P* values as shown in Figure 3. A significant positive correlation was found with the total number of outcome events ($r = 0.305$, $P = 0.01$) and sample size of the included trials ($r = 0.431$, $P = 0.001$). Funding for the trials, journal impact factor, risk of bias domains, and year of publication did not have a significant correlation on regression analysis as shown in Table 5.

DISCUSSION
RCTs are to be interpreted in terms of various factors beyond the reported *P* value which includes sample size, number of events of the outcome, biological plausibility, generalizability, risk of bias involved in the study, conflicts of interest of the authors along with other consideration.[86] Although many studies hail their statistical significance to the 0.05 mark, it has also been a center of controversy for many statisticians.[87] Probable solution to this was laid out with the CIs and Bayesian analyses.[87] We aim not to demerit the significance of *P* value used in all these trials but to propose an additional reporting measure, the FI to augment its interpretation and enhance the validity of the results.

Quantification of fragility of significance was introduced by Feinstein[88] and Walter[89] but was clarified later by Walsh *et al*.[8] Although our study evaluates the FI in spine surgery RCTs, similar studies were performed in various specialties to establish the lack of robustness of the results in the RCTs using FI.[90–92] In our study involving 70 RCTs in spine surgery, the median FI was two, which elaborates on the lack of robustness of the outcomes reported in these RCTs.

Comparison With Previous Work
Our findings compare to the studies on FI in RCTs in fields including critical care[93] (median FI = 1 [IQR 1–3.5]), sports surgery[94] (median FI = 2 [IQR1–2.8]), trauma[95] (median FI = 3 [1–8]) with median sample size 126.6, 64, 168, respectively, which was in line with our median sample size of 133. Most of the studies included in the analysis did not furnish power calculations for sample size and are underpowered particularly when the effect sizes are small. Hence, they report differences occurring out of random chance to be statistically significant and thus are revealed fragile when FI was applied.

TABLE 3. RCTs Included in the Review; n = 70 Trials

Sl. No	Author	Year	Journal of Publication	Total Sample Size	Total events		*P*	FI	FQ
					Control Group	Study Group			
1	Hiller et al[17]	2012	*Spine*	36	3	1	0.01	0	0
2	Arnold et al[18]	2016	*Spine*	319	141	145	0.0004	0	0
3	Peters et al[19]	2015	*Spine*	38	10	2	0.013	2	0.0526
4	He et al[20]	2014	*Spine*	210	0	7	<0.001	6	0.0285
5	Bai et al[21]	2012	*J Spinal Disord Tech*	694	47	15	0.001	19	0.0273
6	Barth et al[22]	2017	*Acta Neurochirurgica*	543	0	60	<0.0001	41	0.0755
7	Strömqvist et al[23]	2010	*Spine*	100	3	13	0.04	2	0.02
8	Bonfill et al[24]	2017	*Spine*	489	4	12	0.04	1	0.0020
9	Klazen et al[25]	2010	*Am J Neuroradiol*	202	35	11	<0.001	15	0.0742
10	Sköld et al[26]	2013	*ESJ*	152	11	30	0.03	5	0.0328
11	Coric et al[27]	2011	*JNS - Spine*	269	82	101	0.05	3	0.0111
12	Murrey et al[28]	2009	*The Spine Journal*	209	8	1	0.033	0	0
13	Baskin et al[29]	2003	*Spine*	33	10	14	<0.05	0	0
14	Dawson et al[30]	2009	*JBJS*	44	12	19	0.05	0	0
15	Delawi et al[31]	2016	*JBJS*	119	44	30	0.03	4	0.0336
16	Xu et al[32]	2017	*JOSR*	80	3	3	0.001	0	0
17	Kallmes et al[33]	2009	*NEJM*	136	7	34	<0.001	14	0.1029
18	Ringel et al[34]	2012	*Spine*	298	93	85	0.019	2	0.0067
19	Phillips et al[35]	2013	*Spine*	403	98	142	0.0001	1	0.0024
20	Garcia et al[36]	2015	*Spine*	324	36	106	<0.01	3	0.0092
21	Glassman et al[9]	2008	*Spine*	102	20	8	0.014	2	0.0196
22	Roh et al[37]	2014	*Spine*	196	66	49	0.014	3	0.0153
23	Lofgren et al[38]	2010	*ESJ*	80	36	27	0.01	2	0.025
24	Han et al[39]	2015	*Zhonghua Yi Xue Za Zhi*	376	260	93	<0.05	NP	NP
25	Hurlbert et al[40]	2013	*Spine*	197	64	89	0.007	16	0.0812
26	Zigler et al[41]	2013	*Spine*	209	11	2	0.0292	2	0.0095
27	Bible et al[42]	2012	*The Spine Journal*	105	9	1	0.016	1	0.0095
28	Wu et al[43]	2014	*J Spinal Disord Tech*	82	24	14	0.046	1	0.0121
29	Burkus et al[44]	2010	*JNS - Spine*	541	221	252	0.006	5	0.0092
30	Thalgott et al[45]	2009	*Spine*	100	1	6	0.026	1	0.01
31	Blasco et al[46]	2012	*JBMR*	125	8	17	0.0462	0	0
32	Dimar et al[47]	2009	*JBJS*	463	151	186	0.014	2	0.0043
33	Nagahama et al[48]	2011	*JNS - Spine*	36	11	18	<0.01	1	0.0277
34	O'Neill et al[49]	2014	*Orthopedics*	40	10	18	0.04	2	0.05
35	Kim et al[50]	2017	*MRCAS*	156	13	0	<0.001	5	0.0320
36	Korovessis et al[51]	2014	*The Spine Journal*	182	86	69	<0.001	11	0.0604
37	Kubota et al[52]	2019	*The Spine Journal*	134	52	60	0.012	6	0.0447
38	Cheng et al[53]	2009	*International Orthopaedics*	97	9	3	0.036	0	0
39	Cheng et al[54]	2011	*CORR*	83	7	1	0.05	0	0
40	Lavelle et al[55]	2019	*Spine*	463	78	93	<0.01	0	0
41	Aglio et al[56]	2014	*JNS - Spine*	58	17	19	<0.05	0	0
42	Liovitz et al[57]	2002	*Spine*	243	43	64	<0.05	6	0.0246
43	Engquist et al[58]	2013	*Spine*	63	20	27	0.01	1	0.0158
44	Coughlan et al[59]	2018	*Spine*	62	12	25	0.04	1	0.0161
45	Kanayama et al[60]	2006	*Spine*	19	9	7	0.05	0	0
46	Merc et al[61]	2013	*Arch Orthop Trauma Surg*	108	21	6	<0.001	5	0.0462
47	Putzier et al[62]	2009	*ESJ*	44	10	3	0.041	1	0.0227
48	Farrokhi et al[63]	2011	*JNS - Spine*	82	1	6	<0.01	0	0
49	Nandyala et al[64]	2014	*Spine*	52	24	17	0.01	1	0.0192
50	Ovadia et al[65]	2018	*Spine*	100	1	6	0.05	0	0

TABLE 3 (*Continued*)

Sl. No	Author	Year	Journal of Publication	Total Sample Size	Total events		*P*	FI	FQ
					Control Group	Study Group			
51	Phillips et al[66]	2015	*Spine*	403	54	48	0.006	0	0
52	Hart et al[67]	2014	*The Spine Journal*	80	32	16	0.003	7	0.0875
53	Rajasekaran et al[68]	2007	*Spine*	478	4	54	<0.001	32	0.0669
54	Rasmussen et al[69]	2008	*Spine*	200	38	20	0.008	5	0.025
55	Sasso et al[70]	2004	*Spine*	139	8	23	<0.001	2	0.0143
56	Sasso et al[71]	2004	*Spine*	140	27	64	<0.001	15	0.1071
57	Ruetten et al[72]	2009	*JNS - Spine*	192	8	0	0.01	1	0.0052
58	Ohtori et al[73]	2011	*ESJ*	82	0	6	0.025	2	0.0243
59	Glassman et al[74]	2007	*Spine*	148	64	75	0.016	2	0.0135
60	Berg et al[75]	2009	*ESJ*	152	11	24	0.031	1	0.0065
61	Thome et al[76]	2018	*The Spine Journal*	554	176	126	<0.001	26	0.0469
62	Jiya et al[77]	2009	*Spine*	26	6	12	0.0302	0	0
63	Pitzen et al[78]	2009	*Spine*	132	4	0	0.045	1	0.0075
64	Jenkins et al[79]	2018	*JBJS*	54	6	2	0.053	0	0
65	Vogl et al[80]	2013	*Spine*	104	34	42	0.0012	NP	NP
66	Gauger et al[81]	2009	*J Paediatr Orthop*	38	17	7	0.007	2	0.0526
67	Wu et al[82]	2010	*Chinese Journal of Traumatology*	176	0	4	<0.001	3	0.0170
68	Chen et al[83]	2013	*ESJ*	80	16	23	0.01	0	0
69	Wu et al[84]	2011	*J Spinal Disord Tech*	677	45	25	0.003	15	0.0221
70	Yang et al[85]	2012	*ESJ*	76	19	25	0.001	0	0

AM J Neuroradiol indicates American Journal of Neuroradiology; Arch Orthop Trauma Surg, Achieves of Orthopaedics and Trauma Surgery; CORR, Clinical Orthopaedics and Related Research; ESJ, European Spine Journal; FI, Fragility Index; FQ, Fragility Quotient; J Paediatr Orthop, Journal of Paediatric Orthopaedics; J Spinal Disord Tech, Journal of Spinal Disorders and Techniques; JBJS, Journal of Bone and Joint Surgery (American); JBMR, Journal of Bone and Mineral Research; JNS, Journal of Neurosurgery; JOSR, Journal of Orthopaedic Surgery and Research; MRCAS, The International Journal of Medical Robotics and Computer Assisted Surgery; NEJM, New England Journal of Medicine; NP, not possible.

In our review we found a significant association between FI and *P* value of the reported outcome, total sample size, the total number of events, and journal impact factor. Previous studies have found similar associations between these factors in mixed combinations.[93–96] This could be due to rounding off of the *P* values before being reported in original trials. Moreover, year of publication did not correlate with FI values which showed a lack of awareness about the issue despite its vast advocacy by various authors of multiple disciplines.

Patient Lost to Follow-up and FI

In a well-designed RCT, sample size calculation could have accounted for 10% of patients lost to follow-up, to retain the validity of the results obtained. 41.8% of the trials in our study had patients lost to follow-up of >10%, which would strongly affect the significance of the results. Moreover, one cannot plan the margin of significance of the results which is based on the event outcomes measured between the two groups at the end of the study. Hence even a well-powered study may suffer from low FI since it also depends on the event outcomes ($r = 0.305$, $P = 0.01$) as shown in our analysis. Moreover, 28 of the 70 included trials had the number of patients lost to follow-up more than the FI which shows that the statistical significance of their results could have been altered by them. Hence along with FI, the number of patients lost to follow-up also remains as a deterministic factor in deciding the validity of the results of even the largest and most rigorously designed trials.

Limitations

The concept of FI has its limitations. It can be applied only to a dichotomous outcome in a 1:1 parallel study design which resulted in the elimination of most of the studies from preliminary screening which could have biased toward a lower median FI in the included trials. We also excluded most of the noninferiority trials since they did not show a statistically significant dichotomous outcome measure. Although a similar statistical component could exist for continuous variables, its objective assessment would be limited by the heterogeneity in the units of measure across the trials.

However, predefining the minimum clinically important difference (MCID)[97] required to report the clinical success of an intervention based on the patient-reported outcome measures (PROM) such as VAS, ODI scores for a given population would let them be converted into a dichotomous outcome which can be utilized for fragility analysis. Only a

TABLE 4. Characteristics of Included RCTs; n = 70 Trials

Characteristic	No. of Studies (%)
Journal name	
Spine	31 (44.3)
European Spine Journal	07 (10)
The Spine Journal	06 (8.6)
Journal of NeuroSurgery - Spine	06 (8.6)
Journal of Bone and Joint Surgery (American)	04 (5.7)
Journal of Spinal Disorders and Techniques	03 (4.3)
Others	13 (18.5)
Outcome used to Calculate Fragility Index	
Primary	44 (62.8)
Secondary	26 (37.1)
Reported *P* values	
<0.05–0.01	38 (54.2)
<0.01–0.001	19 (27.1)
<0.001	13 (18.5)
Funding	
Yes	37 (52.8)
No	26 (37.1)
Not reported	07 (10)
Year of publication	
2002–2005	04 (5.7)
2006–2010	22 (31.4)
2011–2015	32 (45.7)
2016–2019	12 (17.1)

very few RCTs have used a clinical success cut-off for their PROMs to assess the significance of their intervention. Hence, we recommend their inclusion in all RCTs, which utilize PROMs to define their clinical outcomes, during the trial registration process.

Since FI depends largely on the sample size of the trial, FQ was developed which is FI-corrected for the sample size of a trial.[98] However, this FQ decreased the easiness and intuitiveness of an absolute FI; hence, we prefer to use the absolute FI which corresponds directly to the exact number of patients that would have changed the results of the trial.

Implications

Small FI does not imply that an estimated effect is wrong but highlights the fact that changing a small number of events alters the significance of the *P* value over the threshold. Moreover, the applicability of the FI does not devalue the results of the past trials, many of which may never be replicated for various ethical and cost reasons. Meanwhile, FI would serve as an intuitive easily understandable tool used by the researchers and the clinicians to evaluate the robustness of the outcomes of the trials before its clinical application for patient care.

Since FI is a surrogate marker for *P* value, from the correlation curve between *P* value and FI as shown in Figure 3, we chose the minimum value of the 95% CI of the correlation curve as the minimum necessary value of FI. Hence, for a study to be clinically significant rather than having a marginal statistical significance, authors consider

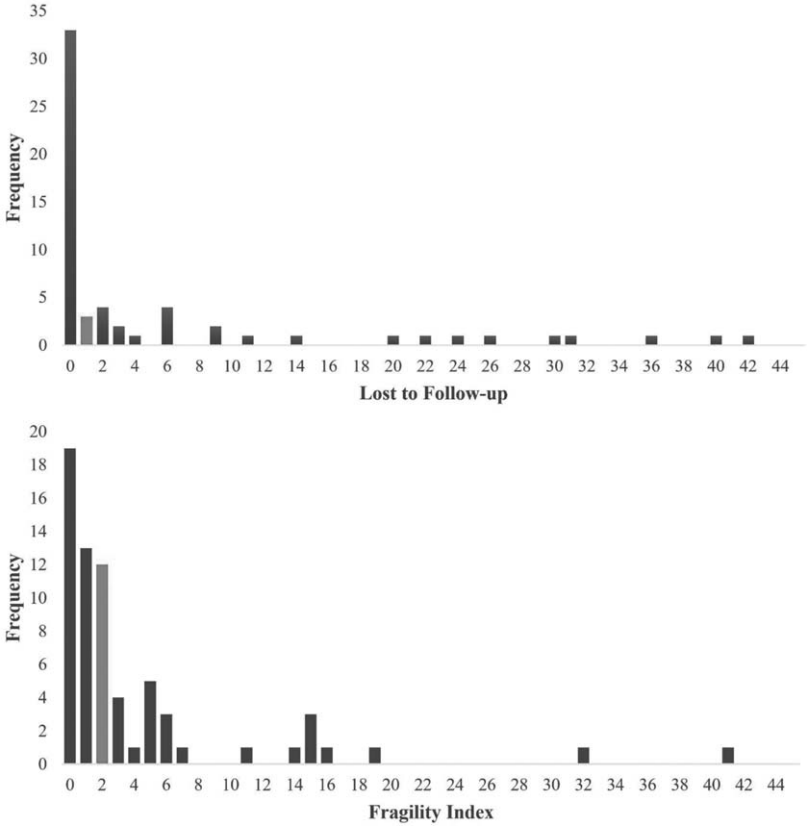

Figure 2. Distribution of Patients lost to follow up and Fragility Index. * Median.

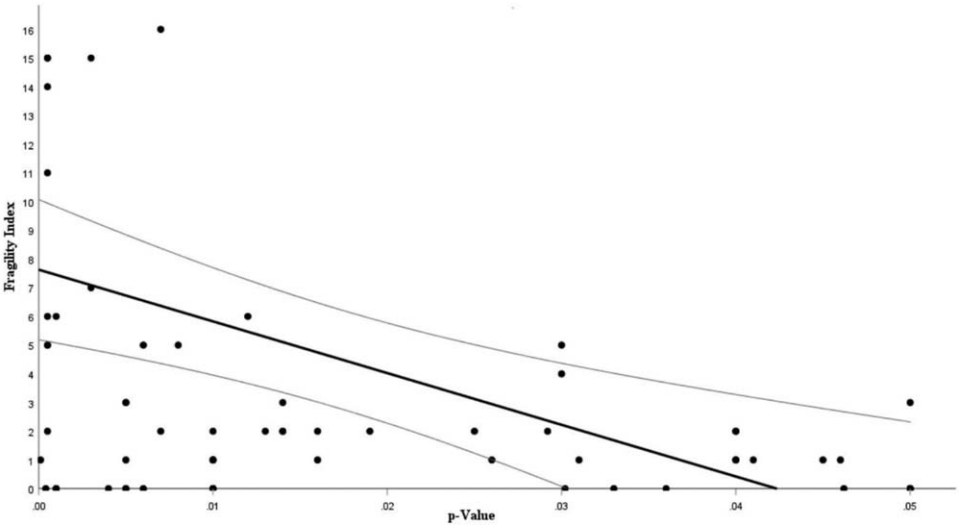

Figure 3. Showing the linear correlation curve between the Fragility Index and the reported *P* values of the included trials along with 95% confidence interval curves.

FI value of ≥5 as an acceptable value to gauge the clinical robustness of its results provided there are no patients lost to follow-up. We hope that the application of FI contributes to the evaluation of the trials holistically in the face of uncertainty of the events that are being tested and can be recommended for routine reporting during publication of the RCTs.

TABLE 5. Fragility Index Across the Subgroups and Their Correlation With Fragility Index.

Characteristic	Median Fragility Index (IQR)	Correlation Coefficient (*P*)
***P* values**		
<0.05–0.01	1 (0–2)	**−0.392 (0.001)**
<0.01–0.001	3 (0–15)	
<0.001	11 (3.5–20.5)	
Impact factor		
<2.5	4 (1.25–18)	0.067 (0.580)
>2.5	1.5 (0–5)	
Sample size		
<100	0 (0–1.5)	**0.431 (0.001)**
100–200	3 (1–5.75)	
>200	3 (1–17)	
No. of events		
0–50	1 (0–2)	**0.305 (0.01)**
50–100	15 (4.5–36.5)	
>100	3 (1.5–8.5)	
Funding		
Yes	2 (0–8.5)	0.049 (0.753)
No	2 (0.5–5)	
Year		
2002–2010	2 (0.75–5)	0.080 (0.511)
2011–2019	2 (0–5.75)	

IQR indicates interquartile range.

CONCLUSION

Overall, statistically significant dichotomous outcomes reported in spine surgery RCTs are more often fragile and the outcomes of the patients lost to follow-up could have changed the significance of the results and hence it needs caution before transcending their results into clinical application for patient care. FI provides an equal and conceptually simple quantification of the robustness of the outcomes in a trial; hence, the addition of FI in routine reporting of RCTs would guide the readers on the robustness of the statistical significance of the outcomes. RCTs with FI ≥5 without any patient lost to follow-up can be considered to have clinically robust results.

➢ Key Points

- ❑ Statistically significant dichotomous outcomes reported in spine surgery RCTs are more often fragile based on FI.
- ❑ The median FI score was two, which means if we switch two patients from nonevent to event, the statistical significance of the outcome is lost.
- ❑ The FI score was less than the number of patients lost to follow-up in 28 of 70 trials.
- ❑ Establishing a clinical success criteria based on MCID makes PROMs eligible for fragility analysis.
- ❑ Results of RCTs with FI ≥5 can be considered clinically robust, provided there are no patients lost to follow-up.

References

1. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am* 2001;83-A:916–26.
2. Bhandari M, Montori VM, Schemitsch EH. The undue influence of significant p-values on the perceived importance of study results. *Acta Orthop* 2005;76:291–5.

3. van Oldenrijk J, van Berkel Y, Kerkhoffs GM, et al. Do authors report surgical expertise in open spine surgery related randomized controlled trials? A systematic review on quality of reporting. *Spine (Phila Pa 1976)* 2013;38:857–64.
4. Bailey CS, Fisher CG, Dvorak MF. Type II error in the spine surgical literature. *Spine (Phila Pa 1976)* 2004;29:1146–9.
5. Sathish M, Eswar R. Systematic reviews and meta-analysis in spine surgery—how good are they in methodological quality? A systematic review. *Global Spine J* 2020;2192568220906810.
6. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 2010;303:1180–7.
7. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203–9.
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;67:622–8.
9. Glassman SD, Carreon LY, Djurasovic M, et al. RhBMP-2 versus iliac crest bone graft for lumbar spine fusion: a randomized, controlled trial in patients over sixty years of age. *Spine (Phila Pa 1976)* 2008;33:2843–9.
10. Bederman SS, Wright JG. Randomized trials in surgery: how far have we come?. *J Bone Joint Surg Am* 2012;94:2–6.
11. Moher D, Liberati A, Tetzlaff J, et al., PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
12. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
13. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–3.
14. Higgins JP, Greene S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. The Cochrane Collaboration: 2011. Available at: www.cochrane-handbook.org Accessed March 28, 2020.
15. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400.
16. Kane SP. Fragility Index Calculator. ClinCalc: available at: https://clincalc.com/Stats/FragilityIndex.aspx. Updated July 19, 2018. Accessed March 28, 2020.
17. Hiller A, Helenius I, Nurmi E, et al. Acetaminophen improves analgesia but does not reduce opioid requirement after major spine surgery in children and adolescents. *Spine (Phila Pa 1976)* 2012;37:E1225–31.
18. Arnold PM, Sasso RC, Janssen ME, et al. Efficacy of i-factor bone graft versus autograft in anterior cervical discectomy and fusion: results of the prospective, randomized, single-blinded food and drug administration investigational device exemption study. *Spine (Phila Pa 1976)* 2016;41:1075–83.
19. Peters A, Verma K, Slobodyanyuk K, et al. Antifibrinolytics reduce blood loss in adult spinal deformity surgery: a prospective, randomized controlled trial. *Spine (Phila Pa 1976)* 2015;40:E443–9.
20. He B, Yan L, Guo H, et al. The difference in superior adjacent segment pathology after lumbar posterolateral fusion by using 2 different pedicle screw insertion techniques in 9-year minimum follow-up. *Spine (Phila Pa 1976)* 2014;39:1093–8.
21. Bai Y-S, Niu Y-F, Chen Z-Q, et al. Comparison of the pedicle screws placement between electronic conductivity device and normal pedicle finder in posterior surgery of scoliosis. *J Spinal Disord Tech* 2013;26:316–20.
22. Barth M, Weiß C, Bouma GJ, et al. Endplate changes after lumbar discectomy with and without implantation of an annular closure device. *Acta Neurochir (Wien)* 2018;160:855–62.
23. Stromqvist BH, Berg S, Gerdhem P, et al. X-stop versus decompressive surgery for lumbar neurogenic intermittent claudication: randomized controlled trial with 2-year follow-up. *Spine (Phila Pa 1976)* 2013;38:1436–42.
24. Bonfill X, Rigau D, Esteban-Fuertes M, et al. Efficacy and safety of urinary catheters with silver alloy coating in patients with spinal cord injury: a multicentric pragmatic randomized controlled trial. The ESCALE trial. *Spine J* 2017;17:1650–7.
25. Klazen CA, Venmans A, de Vries J, et al. Percutaneous vertebroplasty is not a risk factor for new osteoporotic compression fractures: results from VERTOS II. *AJNR Am J Neuroradiol* 2010;31:1447–50.
26. Sköld C, Tropp H, Berg S. Five-year follow-up of total disc replacement compared to fusion: a randomized controlled trial. *Eur Spine J* 2013;22:2288–95.
27. Coric D, Nunley PD, Guyer RD, et al. Prospective, randomized, multicenter study of cervical arthroplasty: 269 patients from the KineflexjC artificial disc investigational device exemption study with a minimum 2-year follow-up: clinical article. *J Neurosurg Spine* 2011;15:348–58.
28. Murrey D, Janssen M, Delamarter R, et al. Results of the prospective, randomized, controlled multicenter food and drug administration investigational device exemption study of the ProDisc-C total disc replacement versus anterior discectomy and fusion for the treatment of 1-level symptomatic cervical disc disease. *Spine J* 2009;9:275–86.
29. Baskin DS, Ryan P, Sonntag V, et al. A prospective, randomized, controlled cervical fusion study using recombinant human bone morphogenetic protein-2 with the CORNERSTONE-SR allograft ring and the ATLANTIS anterior cervical plate. *Spine (Phila Pa 1976)* 2003;28:1219–25.
30. Dawson E, Bae HW, Burkus JK, et al. Recombinant human bone morphogenetic protein-2 on an absorbable collagen sponge with an osteoconductive bulking agent in posterolateral arthrodesis with instrumentation. A prospective randomized trial. *J Bone Joint Surg Am* 2009;91:1604–13.
31. Delawi D, Jacobs W, van Susante JLC, et al. OP-1 compared with iliac crest autograft in instrumented posterolateral fusion: a randomized, multicenter non-inferiority trial. *J Bone Joint Surg Am* 2016;98:441–8.
32. Xu D, Zhuang Q, Li Z, et al. A randomized controlled trial on the effects of collagen sponge and topical tranexamic acid in posterior spinal fusion surgeries. *J Orthop Surg Res* 2017;12:166–1166.
33. Kallmes DF, Comstock BA, Heagerty PJ, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med* 2009;361:569–79.
34. Ringel F, Stüer C, Reinke A, et al. Accuracy of robot-assisted placement of lumbar and sacral pedicle screws: a prospective randomized comparison to conventional freehand screw implantation. *Spine (Phila Pa 1976)* 2012;37:E496–501.
35. Phillips FM, Lee JY, Geisler FH, et al. A prospective, randomized, controlled clinical investigation comparing PCM cervical disc arthroplasty with anterior cervical discectomy and fusion. 2-year results from the US FDA IDE clinical trial. *Spine (Phila Pa 1976)* 2013;38:E907–18.
36. Garcia J, Rolando. Yue JJ, et al. Lumbar total disc replacement for discogenic low back pain: two-year outcomes of the activl multicenter randomized controlled IDE clinical trial. *Spine (Phila Pa 1976)* 2015;40:1873–81.
37. Roh GU, Yang SY, Shim JK, et al. Efficacy of palonosetron versus ramosetron on preventing opioid-based analgesia-related nausea and vomiting after lumbar spinal surgery: a prospective, randomized, and double-blind trial. *Spine (Phila Pa 1976)* 2014;39:E543–9.
38. Lofgren H, Engquist M, Hoffmann P, et al. Clinical and radiological evaluation of trabecular metal and the Smith-Robinson technique in anterior cervical fusion for degenerative disease: a prospective, randomized, controlled study with 2-year follow-up. *Eur Spine J* 2010;19:464–73.
39. Han X, Tian W, Liu B, et al. Necessity of intraoperative three-dimensional navigation for low lumbar spine: a randomized controlled study. *Zhonghua Yi Xue Za Zhi* 2015;95:1796–800.
40. Hurlbert RJ, Alexander D, Bailey S, et al. rhBMP-2 for posterolateral instrumented lumbar fusion: a multicenter prospective

randomized controlled trial. *Spine (Phila Pa 1976)* 2013;38:2139–48.
41. Zigler JE, Delamarter R, Murrey D, et al. ProDisc-C and anterior cervical discectomy and fusion as surgical treatment for single-level cervical symptomatic degenerative disc disease: five-year results of a Food and Drug Administration study. *Spine (Phila Pa 1976)* 2013;38:203–9.
42. Bible JE, O'Neill KR, Crosby CG, et al. Implant contamination during spine surgery. *Spine J* 2013;13:637–40.
43. Wu J, Jin Y, Zhang J, et al. Hemostatic techniques following multilevel posterior lumbar spine surgery: a randomized control trial. *J Spinal Disord Tech* 2014;27:442–6.
44. Burkus JK, Haid RW, Traynelis VC, et al. Long-term clinical and radiographic outcomes of cervical disc replacement with the prestige disc: results from a prospective randomized controlled clinical trial. *J Neurosurg Spine* 2010;13:308–18.
45. Thalgott JS, Fogarty ME, Giuffre JM, et al. A prospective, randomized, blinded, single-site study to evaluate the clinical and radiographic differences between frozen and freeze-dried allograft when used as part of a circumferential anterior lumbar interbody fusion procedure. *Spine (Phila Pa 1976)* 2009;34:1251–6.
46. Blasco J, Martinez-Ferrer A, Macho J, et al. Effect of vertebroplasty on pain relief, quality of life, and the incidence of new vertebral fractures: a 12-month randomized follow-up, controlled trial. *J Bone Miner Res* 2012;27:1159–66.
47. Dimar JR 2nd, Glassman SD, Burkus JK, et al. Clinical and radiographic analysis of an optimized rhBMP-2 formulation as an autograft replacement in posterolateral lumbar spine arthrodesis. *J Bone Joint Surg Am* 2009;91:1377–86.
48. Nagahama K, Kanayama M, Togawa D, et al. Does alendronate disturb the healing process of posterior lumbar interbody fusion? A prospective randomized trial. *J Neurosurg Spine* 2011;14:500–7.
49. O'Neill KR, Lockney DT, Bible JE, et al. Bupivacaine for pain reduction after iliac crest bone graft harvest. *Orthopedics* 2014;37:e428–34.
50. Kim H-J, Jung W-I, Chang B-S, et al. A prospective, randomized, controlled trial of robot-assisted vs freehand pedicle screw fixation in spine surgery. *Int J Med Robot* 2017;13:; 10.1002/rcs.1779.
51. Korovessis P, Vardakastanis K, Repantis T, et al. Balloon kyphoplasty versus KIVA vertebral augmentation—comparison of 2 techniques for osteoporotic vertebral body fractures: a prospective randomized study. *Spine (Phila Pa 1976)* 2013;38:292–9.
52. Kubota G, Kamoda H, Orita S, et al. Platelet-rich plasma enhances bone union in posterolateral lumbar fusion: a prospective randomized controlled trial. *Spine J* 2019;19:e34–40.
53. Cheng L, Nie L, Zhang L. Posterior lumbar interbody fusion versus posterolateral fusion in spondylolisthesis: a prospective controlled study in the Han nationality. *Int Orthop* 2009;33:1043–7.
54. Cheng L, Nie L, Li M, et al. Superiority of the bryan disc prosthesis for cervical myelopathy: a randomized study with 3- year follow up. *Clin Orthop Relat Res* 2011;469:3408–14.
55. Lavelle WF, Riew KD, Levi AD, et al. Ten-year outcomes of cervical disc replacement with the BRYAN cervical disc: results from a prospective, randomized, controlled clinical trial. *Spine (Phila Pa 1976)* 2019;44:601–8.
56. Aglio LS, Abd-El-Barr MM, Orhurhu V, et al. Preemptive analgesia for postoperative pain relief in thoracolumbosacral spine operations: a double-blind, placebo-controlled randomized trial. *J Neurosurg Spine* 2018;29:647–53.
57. Linovitz RJ, Pathria M, Bernhardt M, et al. Combined magnetic fields accelerate and increase spine fusion: a double-blind, randomized, placebo controlled study. *Spine (Phila Pa 1976)* 2002;27:1383–9.
58. Engquist M, Lofgren H, Oberg B, et al. Surgery versus nonsurgical treatment of cervical radiculopathy: a prospective, randomized study comparing surgery plus physiotherapy with physiotherapy alone with a 2-year followup. *Spine (Phila Pa 1976)* 2013;38:1715–22.
59. Coughlan M, Davies M, Mostert AK, et al. A prospective, randomized, multicenter study comparing silicated calcium phosphate versus BMP-2 synthetic bone graft in posterolateral instrumented lumbar fusion for degenerative spinal disorders. *Spine (Phila Pa 1976)* 2018;43:E860–8.
60. Kanayama M, Hashimoto T, Shigenobu K, et al. A prospective randomized study of posterolateral lumbar fusion using osteogenic protein-1 (OP-1) versus local autograft with ceramic bone substitute: emphasis of surgical exploration and histologic assessment. *Spine (Phila Pa 1976)* 2006;31:1067–74.
61. Merc M, Drstvensek I, Vogrin M, et al. A multi-level rapid prototyping drill guide template reduces the perforation risk of pedicle screw placement in the lumbar and sacral spine. *Arch Orthop Trauma Surg* 2013;133:893–9.
62. Putzier M, Strube P, Funk JF, et al. Allogeneic versus autologous cancellous bone in lumbar segmental spondylodesis: a randomized prospective study. *Eur Spine J* 2009;18:687–95.
63. Farrokhi MR, Alibai E, Maghami Z. Randomized controlled trial of percutaneous vertebroplasty versus optimal medical management for the relief of pain and disability in acute osteoporotic vertebral compression fractures. *J Neurosurg Spine* 2011;14:561–9.
64. Nandyala SV, Marquez-Lara A, Fineberg SJ, et al. Prospective, randomized, controlled trial of silicate-substituted calcium phosphate versus rhBMP-2 in a minimally invasive transforaminal lumbar interbody fusion. *Spine (Phila Pa 1976)* 2014;39:185–91.
65. Ovadia D, Drexler M, Kramer M, et al. Closed wound subfascial suction drainage in posterior fusion surgery for adolescent idiopathic scoliosis: a prospective randomized control study. *Spine (Phila Pa 1976)* 2019;44:377–83.
66. Phillips FM, Geisler FH, Gilder KM, et al. Long-term outcomes of the US FDA IDE prospective, randomized controlled clinical trial comparing PCM cervical disc arthroplasty with anterior cervical discectomy and fusion. *Spine (Phila Pa 1976)* 2015;40:674–83.
67. Hart R, Komzák M, Okál F, et al. Allograft alone versus allograft with bone marrow concentrate for the healing of the instrumented posterolateral lumbar fusion. *Spine J* 2014;14:1318–24.
68. Rajasekaran S, Vidyadhara S, Ramesh P, et al. Randomized clinical study to compare the accuracy of navigated and non-navigated thoracic pedicle screws in deformity correction surgeries. *Spine (Phila Pa 1976)* 2007;32:E56–64.
69. Rasmussen S, Krum-Møller DS, Lauridsen LR, et al. Epidural steroid following discectomy for herniated lumbar disc reduces neurological impairment and enhances recovery: a randomized study with two-year follow-up. *Spine (Phila Pa 1976)* 2008;33:2028–33.
70. Sasso RC, Kitchel SH, Dawson EG. A prospective, randomized controlled clinical trial of anterior lumbar interbody fusion using a titanium cylindrical threaded fusion device. *Spine (Phila Pa 1976)* 2004;29:113–22.
71. Sasso WR, Smucker JD, Sasso MP, et al. Long-term clinical outcomes of cervical disc arthroplasty: a prospective, randomized, controlled trial. *Spine (Phila Pa 1976)* 2017;42:209–16.
72. Ruetten S, Komp M, Merk H, et al. Surgical treatment for lumbar lateral recess stenosis with the full-endoscopic interlaminar approach versus conventional microsurgical technique: a prospective, randomized, controlled study. *J Neurosurg Spine* 2009;10:476–85.
73. Ohtori S, Suzuki M, Koshi T, et al. Single-level instrumented posterolateral fusion of the lumbar spine with a local bone graft versus an iliac crest bone graft: a prospective, randomized study with a 2-year follow-up. *Eur Spine J* 2011;20:635–9.
74. Glassman SD, Dimar JR 3rd, Burkus K, et al. The efficacy of rhBMP-2 for posterolateral lumbar fusion in smokers. *Spine (Phila Pa 1976)* 2007;32:1693–8.
75. Berg S, Tullberg T, Branth B, et al. Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. *Eur Spine J* 2009;18:1512–9.
76. Thomé C, Klassen PD, Bouma GJ, et al. Annular closure in lumbar microdiscectomy for prevention of reherniation: a randomized clinical trial. *Spine J* 2018;18:2278–87.
77. Jiya TU, Smit T, van Royen BJ, et al. Posterior lumbar interbody fusion using non resorbable poly-ether-ether-ketone versus resorbable poly-L-lactide-co-D,L-lactide fusion devices. Clinical

outcome at a minimum of 2-year follow-up. *Eur Spine J* 2011;20:618–22.
78. Pitzen TR, Chrobok J, Stulik J, et al. Implant complications, fusion, loss of lordosis, and outcome after anterior cervical plating with dynamic or rigid plates: two-year results of a multi-centric, randomized, controlled study. *Spine (Phila Pa 1976)* 2009;34:641–6.
79. Jenkins TJ, Nair R, Bhatt S, et al. The effect of local versus intravenous corticosteroids on the likelihood of dysphagia and dysphonia following anterior cervical discectomy and fusion: a single-blinded, prospective, randomized controlled trial. *J Bone Joint Surg Am* 2018;100:1461–72.
80. Vogl TJ, Pflugmacher R, Hierholzer J, et al. Cement directed kyphoplasty reduces cement leakage as compared with vertebroplasty: results of a controlled, randomized trial. *Spine (Phila Pa 1976)* 2013;38:1730–6.
81. Gauger VT, Voepel-Lewis TD, Burke CN, et al. Epidural analgesia compared with intravenous analgesia after pediatric posterior spinal fusion. *J Pediatr Orthop* 2009;29:588–93.
82. Wu H, Gao ZL, Wang JC, et al. Pedicle screw placement in the thoracic spine: a randomized comparison study of computer-assisted navigation and conventional techniques. *Chin J Traumatol* 2010;13:201–5.
83. Chen Y, Wang X, Lu X, et al. Comparison of titanium and polyetheretherketone (PEEK) cages in the surgical treatment of multilevel cervical spondylotic myelopathy: a prospective, randomized, control study with over 7-year follow-up. *Eur Spine J* 2013;22:1539–46.
84. Wu Z-X, Huang L-Y, Sang H-X, et al. Accuracy and safety assessment of pedicle screw placement using the rapid prototyping technique in severe congenital scoliosis. *J Spinal Disord Tech* 2011;24:444–50.
85. Yang Z, Tan J, Xu Y, et al. Treatment of MM-associated spinal fracture with percutaneous vertebroplasty (PVP) and chemotherapy. *Eur Spine J* 2012;21:912–9.
86. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ* 2005;330:88.
87. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med* 1999;130:995e1004.
88. Feinstein AR. The unit fragility index: an additional appraisal of "statistical significance" for a contrast of two proportions. *J Clin Epidemiol* 1990;43:201e209.
89. Walter SD. Statistical significance and fragility criteria for assessing a difference of two proportions. *J Clin Epidemiol* 1991; 44:1373e1378.
90. Shen C, Shamsudeen I, Farrokhyar F, et al. Fragility of results in ophthalmology randomized controlled trials: a systematic review. *Ophthalmology* 2018;125:642–8.
91. Mazzinari G, Ball L, Serpa Neto A, et al. The fragility of statistically significant findings in randomised controlled anaesthesiology trials: systematic review of the medical literature. *Br J Anaesth* 2018;120:935–41.
92. Noel CW, McMullen C, Yao C, et al. The fragility of statistically significant findings from randomized trials in head and neck surgery. *Laryngoscope* 2018;128:2094–100.
93. Ridgeon EE, Young PJ, Bellomo R, et al. The Fragility Index in multicenter randomized controlled critical care trials. *Crit Care Med* 2016;44:1278e1284.
94. Khan M, Evaniew N, Gichuru M, et al. The fragility of statistically significant findings from randomized trials in sports surgery. *Am J Sports Med* 2017;45:2164e2170.
95. Tignanelli CJ, Napolitano LM. The Fragility Index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg* 2019;154:74–9.
96. Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine J* 2015;15:2188e2197.
97. Chung AS, Copay AG, Olmscheid N, et al. Minimum clinically important difference: current trends in the spine literature. *Spine (Phila Pa 1976)* 2017;42:1096–105.
98. Ahmed W, Fowler RA, McCredie VA. Does sample size matter when interpreting the fragility index?. *Crit Care Med* 2016; 44:e1142–3.