The Fragility and Reliability of Conclusions of Anesthesia and Critical Care Randomized Trials With Statistically Significant Findings: A Systematic Review

François Grolleau, MD, MPH[1,2]; Gary S. Collins, PhD[3]; Andrei Smarandache, BA, BSc[4]; Romain Pirracchio, MD, PhD[5–7]; Clément Gakuba, MD, PhD[1,8]; Isabelle Boutron, MD, PhD[9–12]; Jason W. Busse, DC, PhD[13]; P. J. Devereaux, MD, PhD[14]; Yannick Le Manach MD, PhD[2]

[1]Department of Anesthesiology and Critical Care Medicine, CHU Caen Côte de Nacre, Caen, France.

[2]Departments of Anesthesia & Clinical Epidemiology and Biostatistics, Michael DeGroote School of Medicine, Faculty of Health Sciences, McMaster University and the Perioperative Research Group, Population Health Research Institute, Hamilton, ON, Canada.

[3]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, University of Oxford, Oxford, United Kingdom.

[4]Michael G. DeGroote School of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada.

[5]Department of Anesthesiology and Critical Care Medicine, Hôpital Européen Georges Pompidou, APHP, Université Paris Descartes, Sorbonne Paris Cite, Paris, France.

[6]Department of Anesthesiology and Perioperative Care, University of California San Francisco, San Francisco General Hospital and Trauma Center, San Francisco, CA.

[7]Department of Biostatistics and Medical Informatics, INSERM U-1153, team ECSTRA, Université Paris Diderot, Sorbonne Paris Cite, Paris, France.

[8]Normandie Univ, UNICAEN, INSERM, INSERM UMR-S U1237, "Physiopathology and Imaging of Neurological Disorders" PhIND, Caen, France.

[9]Université Paris Descartes, Sorbonne Paris Cité, Paris, France.

[10]Centre de Recherche Epidémiologie et Statistique, INSERM U1153, Paris, France.

[11]Cochrane France, Paris, France.

[12]Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique-Hôpitaux de Paris, Paris, France.

[13]Departments of Anesthesia and Health Research Methods, Evidence & Impact, and the Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, ON, Canada.

[14]Departments of Medicine & Health Evidence and Impact, Michael DeGroote School of Medicine, Faculty of Health Sciences, McMaster University and the Perioperative Research Group, Population Health Research Institute, Hamilton, ON, Canada.

Drs. Grolleau and Le Manach had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis and contributed to study concept and design and analysis and interpretation of data. Drs. Grolleau and Smarandache contributed to acquisition of data. Dr. Grolleau contributed in drafting of the article. Drs. Grolleau, Collins, Pirracchio, Gakuba, Boutron, Busse, Devereaux, and Le Manach contributed in critical revision of the article for important intellectual content.

DOI: 10.1097/CCM.0000000000003527

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccmjournal).

Dr. Collins was supported by the National Institute for Health Research Biomedical Research Centre, Oxford, United Kingdom. Dr. Devereaux's institution received funding from Abbott Diagnostics, Boehringer-Ingelheim, Covidien, Octopharma, Philips Healthcare, Roche Diagnostics, and Stryker. The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: yannick.lemanach@phri.ca

Objectives: The Fragility Index, which represents the number of patients responsible for a statistically significant finding, has been suggested as an aid for interpreting the robustness of results from clinical trials. A small Fragility Index indicates that the statistical significance of a trial depends on only a few events. Our objectives were to calculate the Fragility Index of statistically significant results from randomized controlled trials of anesthesia and critical care interventions and to determine the frequency of distorted presentation of results or "spin".

Data Sources: We systematically searched MEDLINE from January 01, 2007, to February 22, 2017, to identify randomized controlled trials exploring the effect of critical care medicine or anesthesia interventions.

Study Selection: Studies were included if they randomized patients 1:1 into two parallel arms and reported at least one statistically significant ($p < 0.05$) binary outcome (primary or secondary).

Data Extraction: Two reviewers independently assessed eligibility and extracted data. The Fragility Index was determined for the chosen outcome. We assessed the level of spin in negative trials and the presence of recommendations for clinical practice in positive trials.

Data Synthesis: We identified 166 eligible randomized controlled trials with a median sample size of 207 patients (interquartile

range, 109–497). The median Fragility Index was 3 (interquartile range, 1–7), which means that adding three events to one of the trials treatment arms eliminated its statistical significance. High spin was identified in 42% ($n = 30$) of negative randomized controlled trials, whereas 21% ($n = 20$) of positive randomized controlled trials provided recommendations. Lower levels of spin and recommendations were associated with publication in journals with high impact factors ($p < 0.001$ for both).

Conclusions: Statistically significant results in anesthesia and critical care randomized controlled trials are often fragile, and study conclusions are frequently affected by spin. Routine calculation of the Fragility Index in medical literature may allow for better understanding of trials and therefore enhance the quality of reporting. (*Crit Care Med* 2018; XX:00–00)

Key Words: anesthesia; critical care; intensive care; randomized controlled trials; research methodology; research report

In 2016, the American Statistical Association encouraged researchers to move toward a post *p* values era (1). The Fragility Index (FI) may be a part of the solution by targeting clinicians rather than statisticians. Although *p* values are used to declare statistical significance, the FI represents the number of patients responsible for the statistically significance of a trial finding (2). A large FI indicates that many patients are responsible for the statistically significance of a trial finding, increasing the confidence in the observed treatment effect. Conversely, a low FI indicates that the statistical significance of a trial finding relies on only a few patients; it proves the result fragile, witnessing it might not be reproduced in further studies. The FI is calculated by changing the status of patients without an event to an event in the treatment group with the smallest number of events, until the *p* value exceeded 0.05. Previous systematic reviews have reported high FIs in heart failure trials (3), but low FIs in spine surgery (4), critical care (5), and sport surgery (6) trials. Such findings shed light on how unlikely trial results from specific fields are to be reproduced.

To inflate their findings, possibly to improve chances of publication, authors often consciously or subconsciously over interpret the results of their randomized controlled trial (RCT), that is, they add "spin" to the conclusions of a scientific report. Spin is defined as a nonneutral way of reporting that distorts the interpretation of results and misleads readers (7). Article conclusions, which may include clinical recommendations, are the sections most often affected by spin (7), and they may influence clinical practice in a misleading way.

Our primary aim was to evaluate the robustness of statistically significant results from RCTs assessing anesthesia and critical care interventions by determining their FI. Because most trials in anesthesia and critical care are small, single-center trials (8), we hypothesized that statistically significant findings would often rely on a small number of patients. We also sought to determine the nature and frequency of spin in trial conclusions and its association with FI values.

METHODS

This systematic review was registered (PROSPERO registration number CRD42017057526) and is reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (9). The data supporting this study are freely available via our website https://www.researchgate.net/publication/320559681.

Sources of Information and Search Strategy

We searched MEDLINE from January 1, 2007, to February 22, 2017, to identify all RCTs related to critical care medicine or anesthesia, in any language, that were published in any of 19 anesthesiology journals or 13 critical care journals with the highest impact factors, or six major general journals (*The New England Journal of Medicine, Lancet, JAMA, JAMA Internal Medicine, British Medical Journal, Annals of Internal Medicine).* We used medical subject headings, in various combinations, supplemented with free text to increase sensitivity and the Cochrane sensitivity- and precision-maximizing search strategy for identifying RCTs (10). The full search strategy is presented in the **Supplementary Material 1** (Supplemental Digital Content 1, http://links.lww.com/CCM/E112).

Eligibility Criteria

Eligible studies were 1) superiority trials, 2) randomized patients 1:1 into two parallel arms, 3) reported in their abstract at least one statistically significant binary outcome (i.e., a *p* value of < 0.05 or a 95% CI that excluded a null value), and 4) examined an intervention in the field of anesthesia or critical care. We did not include studies that were systematic reviews, meta-analyses, post hoc analyses, noninferiority trials, and RCTs with cluster or cross-over design.

Study Selection

Two reviewers (F.G., A.S.) independently screened the titles and abstracts and reviewed the full text of all potentially eligible citations using pilot-tested forms. All discrepancies were resolved by consensus.

Data Collection Process

Using a piloted electronic data extraction form, a pair of reviewers (F.G., A.S.) extracted data independently and in duplicate. All extracted data were verified by another reviewer (Y.L.M.). For each eligible RCT, we extracted data for the primary outcome (i.e., number of events and nonevents for each arm, *p* value, and corresponding statistical test). When the primary outcome was not binary or was binary but nonstatistically significant, we also extracted data for one statistically significant secondary binary outcome that was identified in the abstract. When multiple statistically significant secondary binary outcomes were reported in the abstract, we extracted the one considered to be the most patient-important based on a consensus between two clinicians (F.G., Y.L.M.).

We also extracted the following data from each study: journal name, 2015 Journal Impact Factor, year of publication, and funding source (i.e., for-profit, nonprofit, or both; not

reported, no funding). Loss to follow-up for each group was extracted only for articles reporting a Consolidated Standards of Reporting Trials (CONSORT) flow diagram (11). Likewise, we counted articles not reporting any flow diagram and articles reporting a non-CONSORT flow diagram.

Primary outcomes were those explicitly reported as such in the published article. If none was explicitly reported, we considered the outcome used to calculate the sample size; if outcomes were not stated in sample size calculation, we took the outcome as defined in the primary study objective, if available. If the primary outcome was still not clearly identified, the article was excluded.

Trial Registration and Posting of Results

For each included study, two authors (F.G., A.S.) screened full-texts for registration. Studies were classified as: 1) registered on ClinicalTrial.gov, 2) registered on a non-American registry, or 3) not registered. For each registered trial, we checked whether the results were posted on trial registries. We also compared registered and reported primary outcomes, anticipated and actual sample size. Sample size was considered inconsistent when it did not reach 80% of the registered planned sample size. Data extraction from the registries was performed independently and in duplicate by two reviewers (F.G., A.S.) using a piloted electronic form and verified by a third reviewer (Y.L.M.).

Assessment of Conclusions Sections

When the primary outcome was statistically significant, the article conclusion provided by the authors was classified as providing recommendations for clinical practice or not. We did not assess for the presence of spin when primary outcome was significant.

When the primary outcome was not statistically significant, the level of spin in conclusion section was classified as high, moderate, low/absent according to the classification described by Boutron et al (7). "High spin" was defined as no uncertainty in the framing, no recommendation for further trials, and no acknowledgment of the absence of statistically significant results. In addition, when the conclusion section reported recommendations on the use of the treatment in clinical practice, we classified this section as having a high level of spin. "Moderate spin" was defined as some uncertainty in the framing or recommendations for further trials but no acknowledgment of the absence of statistically significant results on the primary outcome. "Low or no spin" was defined as uncertainty in the framing and recommendations for further trials or the acknowledgment of the absence of statistically significant results on the primary outcome.

FI Calculation

We calculated the FI for significant binary outcomes collected according to the method described by Walsh et al (2) using 2×2 contingency tables. Specifically, we recalculated the *p* value using a two-sided Fisher exact test and then added events to the group with the smallest number of events, while subtracting nonevents from the same group to keep the total number of participants constant, until the *p* value reached or exceeded 0.05. The smallest number of additional events required to obtain a *p* value of greater than or equal to 0.05 represented the FI for that outcome.

Statistical Analysis

We reported the median and interquartile range (IQR) of continuous variables, and the number of occurrences with proportions represented as percentages for categorical variables. We used the Kruskal-Wallis test to compare the FI distributions between articles with different levels of spin and between articles with or without clinical recommendations provided in their conclusion. The Pearson's correlation coefficient was used to evaluate correlation between FI and sample size and between FI and *p* values. Sample size and *p* values were log-transformed in order to approximate normal distribution. Preplanned subgroup analyses were carried out for studies published in anesthesia, critical care, and general journals

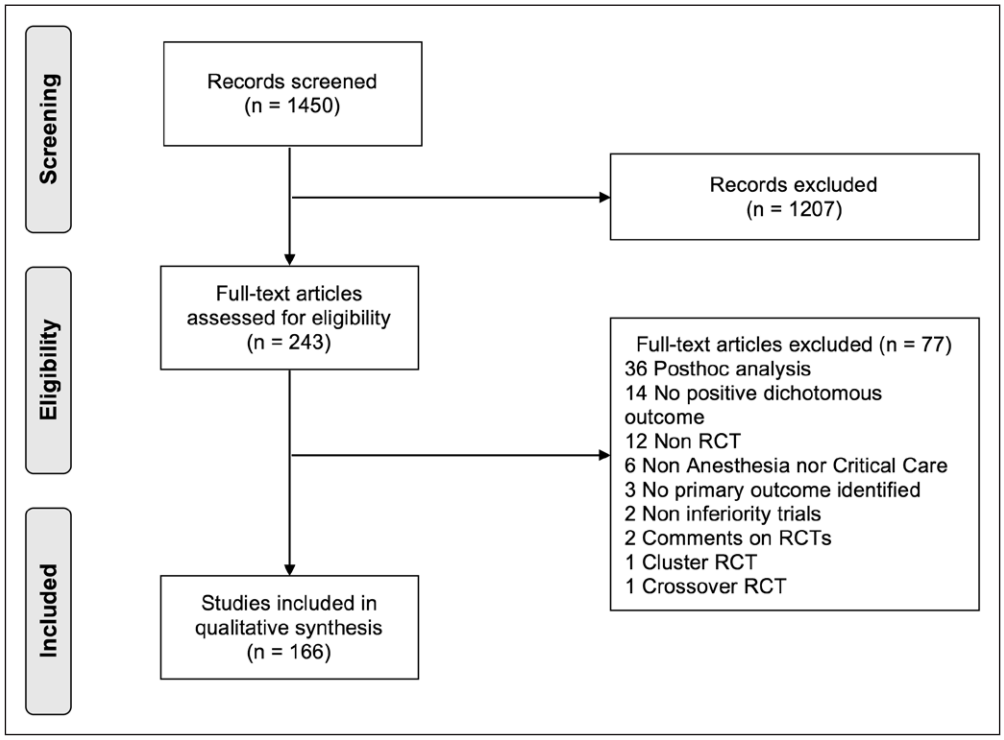

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2009 flow diagram. RCT = randomized controlled trial.

All analyses were performed using R 3.4.2 (http://www.R-project.org; the R Foundation for Statistical Computing, Vienna, Austria). All tests of significance were two-tailed, and *p* value of less than 0.05 was considered significant.

RESULTS

Study Characteristics

Of 1,450 unique citations, 166 reports (11%) proved eligible (**Fig. 1**; and **Supplementary Material 2**, Supplemental Digital Content 2, http://links.lww.com/CCM/E113). The median sample size was 207 (IQR, 109–497), the median number of events was 48 (IQR, 26–99), and the median number of patients lost to follow-up was 4 (IQR, 0–18). A total of 23% of trials ($n = 38$) were unregistered. Among the 128 registered trials, there was a discrepancy between the primary outcome reported in the article and the one appearing in the trial registry in 22 trials (17%). The actual sample size was less than 80% of the calculated sample size in 43 trials (26%). Fifty-three articles (32%) did not present a CONSORT flow diagram, and funding sources were not reported in 31 articles (19%). One-hundred forty-eight trials (89%) had no results posted on registries at the time of the search (**Table 1**).

FI

The median FI for the 166 evaluated outcomes was 3 (IQR, 1–7) (**Fig. 2**). The median FI was the same between studies with a statistically significant primary outcome ($n = 73$; 44%) and studies with a statistically significant secondary outcome ($n = 93$; 56%) (3 [IQR, 1–6] and 3 [1–7], respectively). Furthermore, 21 trails (13%) had a FI of zero as the statistically significant outcome was found nonsignificant when recalculating the *p* value using a two-sided Fisher exact test.

The FI was significantly correlated with *p* values and sample size: as FI values increased, corresponding reported *p* values decreased (after log transformation of *p* values: $r = –0.61$; 95% CI, –0.70 to –0.51; $p < 0.001$) and sample size increased (after log transformation of sample size: $r = 0.40$; 95% CI, 0.27–0.52; $p < 0.001$) (**Fig. 3**). Explanatory subgroup analysis did not show any significant difference in the FI distribution (**Supplementary Material 3**, Supplemental Digital Content 3, http://links.lww.com/CCM/E114).

Level of Spin and Recommendations for Clinical Practice in Conclusions

The level of spin and recommendations for clinical practice in conclusions sections are presented in **Table 2**, and illustrative examples (16–20) are presented in **Supplementary Material 4** (Supplemental Digital Content 4, http://links.lww.com/CCM/E115). We identified high spin in 42% of trials (30/71) with a nonstatistically significant primary outcome. Twenty-one percent of main text conclusion sections of trials (20/95) with a statistically significant primary outcome provided recommendations for clinical practice.

Neither recommendations for clinical practice nor level of spin was associated with FI values (median FI = 3 [IQR, 2–7] vs 3 [IQR, 1–6]; $p = 0.26$ and 2 [IQR, 1–7] vs 4 [IQR, 1–8] vs 3 [IQR, 2–4]; $p = 0.51$ respectively); however, they both were associated with journal impact factor. Journals with higher impact factor published more recommendations for clinical practice in studies with a statistically significant primary outcome (median

TABLE 1. Characteristics of Included Studies ($n = 166$)

Characteristics	No. of Studies, *n* (%)
Type of primary outcome	
Binary	125 (75)
Continuous	41 (25)
Registration	
ClinicalTrial.gov	100 (60)
Non-American registry	28 (17)
No Registration found	38 (23)
Same registered and reported primary outcome	
Yes	97 (58)
No	22 (13)
Unknown	47 (28)
Planned and obtained sample size consistency[a]	
Yes	108 (65)
No	43 (26)
Unknown	15 (9)
Flow diagram	
CONSORT	113 (68)
Non-CONSORT	28 (17)
Not reported	25 (15)
Posting of results on registries	
Yes	18 (11)
No	148 (89)
Funding source	
Profit	22 (13)
Non Profit	78 (47)
Both	32 (19)
No funding	3 (2)
Not reported	31 (19)
Journals	
Anesthesia journals	34 (20)
Critical Care journals	71 (43)
General journals	61 (37)

[a]Defined by an obtained sample size of more than 80% of the planned sample size.

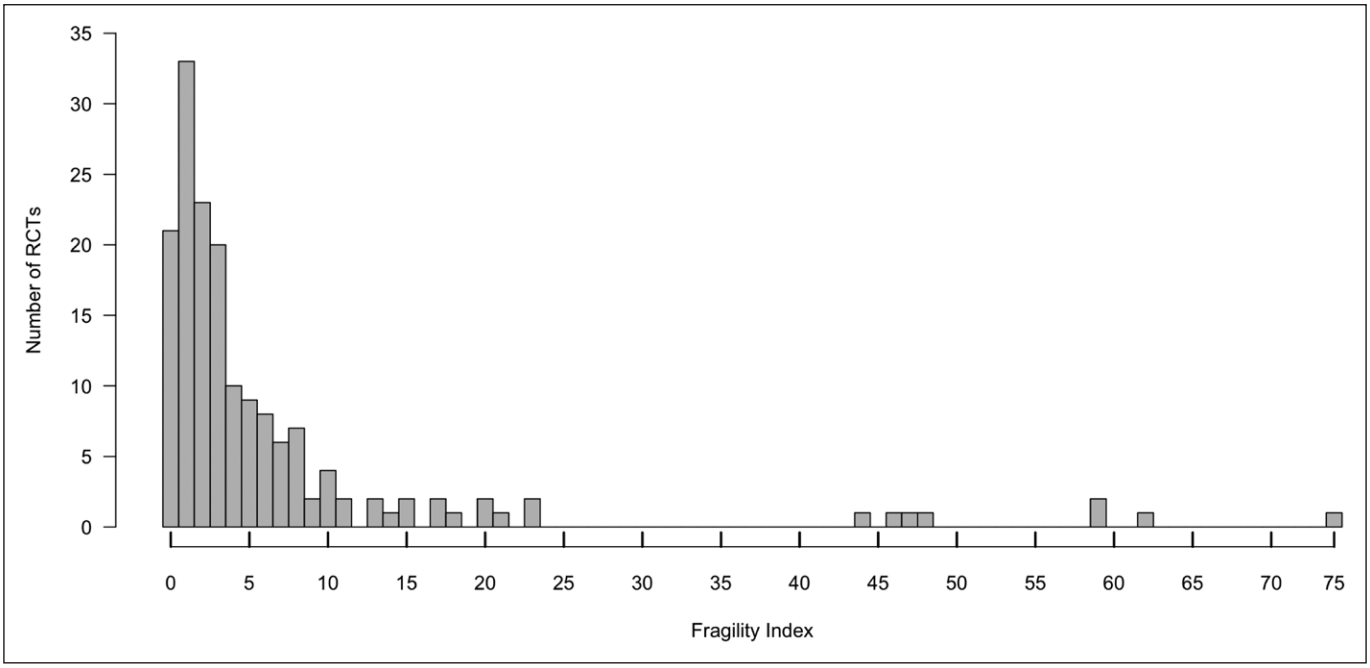

Figure 2. Distribution of Fragility Index from 166 studies. The median Fragility Index was 3 (interquartile range, 1–7). RCT = randomized controlled trial

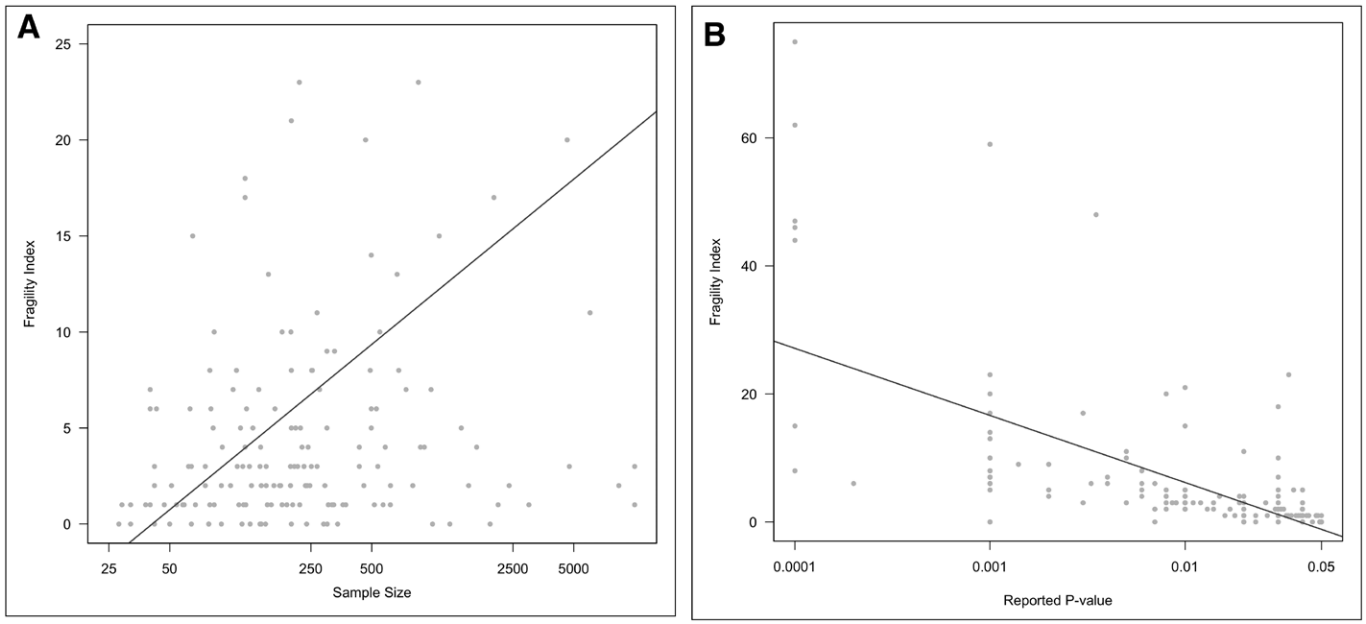

Figure 3. Fragility Index correlations. **A**, Increasing Fragility Index values correlated with increasing sample size (after log transformation of sample size: $r = 0.40$; $p < 0.001$). **B**, Increasing Fragility Index values correlated with decreasing p values (after log transformation of p values: $r = -0.61$; $p < 0.001$).

Impact Factor = 38 [IQR, 9–44] vs 7 [IQR, 5–38]; $p < 0.01$) and had a lower level of spin in studies with no statistically significant primary outcome (median Impact Factor = 38 [IQR, 17–60] vs 7 [IQR, 5–27] vs 7 [IQR, 6–13]; $p < 0.01$).

DISCUSSION

Across 166 anesthesia and critical care trials, we found a median FI of 3 (IQR, 1–7). This means that for 50% of trials, reversing the outcome status of three or less patients in one treatment group would change the interpretation of a trial from a positive to a negative finding. Furthermore, in 13% of articles from our sample, solely p value recalculation using Fisher exact test led to a loss of statistical significance. In addition, we identified a high or moderate level of spin in 69% of trials (49/71) with a nonstatistically significant primary outcome.

Strengths and Limitations

Our sample of studies was not exhaustive since we used fragmented keywords for our search strategy. However, unlike

TABLE 2. Characteristics of Conclusions Sections and Association With Fragility Index and Impact Factors

Conclusions	No. of Studies, *n* (%)	Fragility Index, Median (IQR)	*p*[a]	Impact Factor, Median (IQR)	*p*[a]
Studies with a statistically significant primary outcome	*n* = 95	3 (1–6)			
Recommendations for clinical practice	20 (21)	3 (2–7)	0.26	38 (9–44)	< 0.01
No recommendations for clinical practice	75 (79)	3 (1–6)		7 (5–38)	
Studies with a nonstatistically significant primary outcome	*n* = 71	3 (1–8)			
High spin	30 (42)	2 (1–7)	0.51	7 (6–13)	< 0.01
Moderate spin	19 (27)	4 (1–8)		7 (5–27)	
Low spin/none	22 (31)	3 (2–4)		38 (17–60)	

IQR = interquartile range.
[a]From the Kruskal-Wallis test.

meta-analysis, our review did not intend to estimate any treatment effect. Instead, our goal was to describe a contemporary sample of anesthesia and critical care research. We believe it is very unlikely that including more studies would have changed our estimation in FI distribution. Although our search strategy was not exhaustive for all critical care and anesthesiology trials, it is unlikely that the FI would vary for trials identified by our search strategy compared with trials that were not identified by our search strategy.

Another limitation comes from the FI itself, which can only be calculated for statistically significant binary outcomes. As such, trials reporting only statistically significant continuous outcomes and studies with no statistically significant results were not included.

Last, the assessment of spin is subjective because the strategies used for spin were highly variable and interpretation depended on the context. We increased the reliability of this assessment by having two reviewers extract the data independently using piloted data extraction forms, with any disagreements resolved by consensus.

Relation to Other Studies

Our results are consistent with those found in spine surgery (4), critical care (5), and sport surgery (6) where median FI were 2 (IQR, 1–3), 2 (IQR, 1–3.5), and 2 (IQR, 1–3), respectively. In contrast, heart failure trials had a higher median FI of 26 (IQR, 8.5–39) (3). The prior study of critical care trials was restricted to mortality outcomes and only considered 56 RCTs. Hence, our findings extend results to other outcomes and considered a larger sample of trials.

Anesthesia and critical care trials exhibit unique features. Regardless of the allocated treatment, the occurrence of an outcome such as death may be highly influenced by unexpected rare complications such as peroperative vascular wound or ventilator-associated pneumonia. Thus, balancing prognostic factors between groups may require larger sample sizes than in other medical fields (12, 13).

Consistent with previous studies, FIs correlated strongly with sample size (2, 4, 5), indicating that increasing the number of participants would increase the robustness of anesthesia and critical care findings. Although the FI correlated strongly with *p* values, consistent with earlier studies (2, 4–6, 14), and both metrics have similar mathematical characteristics (14), the FI is more convenient and straightforward for clinicians as it describes uncertainty in terms of number of patients responsible for the positivity of a trial.

With respect to spin, we found it almost twice as prevalent among anesthesia and critical care publications than in other specialties (7).

Implications for Anesthesia and Critical Care Research

Spin is a specific type of reporting bias that has been described as a collective failure involving authors, editors, as well as educators. Contrary to our hypothesis, we found that spin was not associated with high FI values for secondary outcomes; this suggest that for articles with nonstatistically significant primary outcomes, strong conclusions are often not supported by any data. We demonstrated, however, that the higher the journal impact factor, the less likely spin was to occur. Many clinicians do not have the time to read an entire article, as a result they often read only conclusion sections. Conclusions sections should therefore be as reliable as possible and reflect the study findings. Providing FI in research articles may serve to educate the scientific community regarding statistical robustness of findings, reduce the occurrence of spin, and enhance the quality of reporting.

Beyond spin, reporting was frequently incomplete: almost half of the articles eligible for our review did not present a CONSORT flow diagram and funding sources were not reported in approximately one of five articles. We showed that within our sample, nearly one trial out of four was not registered. Although posting of results is mandatory for clinical trials of Food and Drug Administration approved drug and devices, we found 89% of trials did not post their results on registries within the time frame of our search. Reporting was also often inconsistent: articles reported and registered primary outcomes were different in approximately one of five

trials while sample size was much lower than anticipated in one of four trials. Such findings emphasize that anesthesia and critical care researchers alongside with journal editors should work together to decrease this waste in research. In this respect, recent guidelines (11) that target quality and transparency of research (15) must be more widely embraced.

In conclusion, in this representative sample of anesthesia and critical care randomized trials, results were often fragile, and conclusions were frequently inconsistent with the results. We believe that routine calculation of the FI will facilitate the interpretation of uncertainty in scientific reports and may lower the incidence of spin.

REFERENCES

1. Wasserstein RL, Lazar NA: The ASA's statement on p-values: Context, process, and purpose. *Am Stat* 2016; 70:129–133
2. 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; 67:622–628
3. Docherty KF, Campbell RT, Jhund PS, et al: How robust are clinical trials in heart failure? *Eur Heart J* 2017; 38:338–345
4. 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; 15:2188–2197
5. Ridgeon EE, Young PJ, Bellomo R, et al: The Fragility Index in multicenter randomized controlled critical care trials. *Crit Care Med* 2016; 44:1278–1284
6. Khan M, Evaniew N, Gichuru M, et al: The Fragility of statistically significant findings from randomized trials in sports surgery: A systematic survey. *Am J Sports Med* 2017; 45:2164–2170
7. Boutron I, Dutton S, Ravaud P, et al: Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010; 303:2058–2064
8. Sivakumar H, Peyton PJ: Poor agreement in significant findings between meta-analyses and subsequent large randomized trials in perioperative medicine. *Br J Anaesth* 2016; 117:431–441
9. Moher D, Liberati A, Tetzlaff J, et al; PRISMA Group: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *J Clin Epidemiol* 2009; 62:1006–1012
10. Lefebvre C, Manheimer E, Glanville J: Chapter 6: Searching for Studies 2011. Available at: www.handbook.cochrane.org. Accessed February 11, 2017
11. Schulz KF, Altman DG, Moher D; CONSORT Group: CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010; 63:834–840
12. Nguyen TL, Collins GS, Lamy A, et al: Simple randomization did not protect against bias in smaller trials. *J Clin Epidemiol* 2017; 84:105–113
13. Harhay MO, Wagner J, Ratcliffe SJ, et al: Outcomes and statistical power in adult critical care randomized trials. *Am J Respir Crit Care Med* 2014; 189:1469–1478
14. Carter RE, McKie PM, Storlie CB: The Fragility Index: A P-value in sheep's clothing? *Eur Heart J* 2017; 38:346–348
15. Chalmers I, Bracken MB, Djulbegovic B, et al: How to increase value and reduce waste when research priorities are set. *Lancet* 2014; 383:156–165
16. Legriel S, Lemiale V, Schenck M, et al; HYBERNATUS Study Group: Hypothermia for neuroprotection in convulsive status epilepticus. *N Engl J Med* 2016; 375:2457–2467
17. Cruz DN, Antonelli M, Fumagalli R, et al: Early use of polymyxin B hemoperfusion in abdominal septic shock: The EUPHAS randomized controlled trial. *JAMA* 2009; 301:2445–2452
18. Mikor A, Trásy D, Németh MF, et al: Continuous central venous oxygen saturation assisted intraoperative hemodynamic management during major abdominal surgery: A randomized, controlled trial. *BMC Anesthesiol* 2015; 15:82
19. Ferrer M, Sellarés J, Valencia M, et al: Non-invasive ventilvaation after extubation in hypercapnic patients with chronic respiratory disorders: Randomised controlled trial. *Lancet* 2009; 374:1082–1088
20. Olper L, Bignami E, Di Prima AL, et al: Continuous positive airway pressure versus oxygen therapy in the cardiac surgical ward: A randomized trial. *J Cardiothorac Vasc Anesth* 2017; 31:115–121