JAMA Surgery | Review

The Fragility Index in Randomized Clinical Trials as a Means of Optimizing Patient Care

Christopher J. Tignanelli, MD; Lena M. Napolitano, MD

IMPORTANCE The Fragility Index (FI) is the minimum number of participants in a randomized clinical trial (RCT) whose status would have to change from a nonevent (not experiencing the primary end point) to an event (experiencing the primary end point) required to turn a statistically significant result to a nonsignificant result. The FI measures the robustness (or fragility) of the results of an RCT and is an important aid to the clinician's interpretation of RCT results. It has now been recognized that RCTs, which provide the foundation for treatment guideline recommendations, may not be robust.

OBSERVATIONS Most RCTs in surgery and general medicine are fragile (with a low FI score), in contrast to those in cardiac disease and heart failure, where most RCTs are robust (with high FI scores). For clinical trials of trauma, we identified that the median (interquartile range) FI score was 3 (1-8), which means that adding 3 events to the opposite treatment arm in a given RCT eliminated statistical significance. The median Fragility Quotient (the FI score divided by the total study sample size) was 0.016 (0.0043-0.0408).

CONCLUSIONS AND RELEVANCE The provision of high-quality, evidence-based clinical care in surgery for optimal patient outcomes requires a foundation of robust clinical research evidence, and knowledge of the FI will assist in future surgical RCT design. We strongly recommend the routine reporting of FI scores for all future trauma and surgical RCTs to assist in appropriate and optimal decision making in the care of patients who have experienced trauma and/or need surgery. We also recommend the routine inclusion of the FI score in the development of clinical guidelines to assist the clinician in ascertaining whether guideline recommendations are robust. Surgeons should be aware to particularly exercise caution when considering a potential change in clinical practice based on RCTs with a low FI score.

JAMA Surg. 2019;154(1):74-79. doi:10.1001/jamasurg.2018.4318
Published online November 7, 2018.

Author Affiliations: Division of Acute Care Surgery, Department of Surgery, University of Minnesota, Minneapolis (Tignanelli); Division of Acute Care Surgery, Department of Surgery, University of Michigan, Ann Arbor (Napolitano).

Corresponding Author: Lena M. Napolitano, MD, MCCM, Division of Acute Care Surgery Department of Surgery, University of Michigan Health System, University Hospital, 1500 E Medical Dr, Room 1C340A-UH, Ann Arbor, MI 48109-0033 (lenan@umich.edu).

Randomized clinical trials (RCTs) serve as the gold standard and represent the highest level of evidence for the determination of optimal and effective treatment strategies in evidence-based medicine for patients, particularly in terms of statistical reliability. Methods used in RCTs decrease selection bias and minimize confounding, which promote their ability to establish causation.

Clinicians evaluating RCT results in an effort to provide optimal patient care frequently rely on rejection of the null hypothesis, with a *P* value threshold of less than .05 and appropriate 95% CIs used to determine statistical significance and confirm a positive outcome. These positive findings from RCTs frequently lead to substantial changes in clinical practice and patient care. Yet methodologists have noted that medical research studies should not be interpreted from *P* values alone,[1-4] noting that a "high rate of non-replication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05."[2(pe124)]

The Fragility Index (FI) is an important aid in the interpretation of clinical trial results.[5] The FI aids in determination of when statistical significance in a clinical trial may be lost as a result of a shift of a few additional events from the experimental group to the control group.[6] No matter what threshold *P* value is chosen, the FI score will still serve as an additional metric to demonstrate how easily statistical significance may be exceeded. Use of the FI can aid physicians in identifying trials that may be at risk of being overturned by future studies and avoiding overestimating the significance of RCT results. This narrative review discusses the definition and calculation of the FI and Fragility Quotient (FQ) scores, use in the interpretation of medical and surgical RCTs, a detailed review of FI scores in trauma RCTs, and discussion of the importance of this statistic in the development of national clinical guidelines. It provides a framework for surgeons to use the FI in the assessment of RCTs to ultimately provide optimal surgical patient care.

Definition of the FI and FQ

The FI measures the robustness (or fragility) of the results of a clinical trial. The FI is the minimum number of patients whose status would have to change from a nonevent (not experiencing the pri-

mary end point) to an event (experiencing the primary end point) to make the study lose statistical significance. For instance, an FI score of 1 means that only 1 patient would have to not experience the primary end point to make the trial result nonsignificant. In other words, it is a measure of how many events the statistical significance of a clinical trial result depend on. A smaller FI score indicates a more fragile, less statistically robust clinical trial result.

The FQ provides a method to evaluate fragility relative to the sample size of a trial. A smaller FQ also indicates a less robust study outcome.

Calculation of FI and FQ Scores

It is possible to calculate the FI score by converting 1 patient in a control or experimental group from a nonevent outcome to an event outcome and then recalculating a 2-sided Fisher exact test until the *P* value is .05 or greater. This determines how many patients would have to have a different outcome to make a study's results nonsignificant.

An example of FI calculation is provided in **Table 1**. In the Prophylaxis for Thromboembolism in Critical Care Trial (PROTECT) trial (NCT00182143), which compared dalteparin with unfractionated heparin for venous thromboembolism prophylaxis in critically ill patients, pulmonary embolism was confirmed in 24 patients who received dalteparin vs 43 patients who received unfractionated heparin (hazard ratio, 0.51 [95% CI, 0.30-0.88]; *P* = .01). The FI score for the PROTECT trial is 2; in other words, if 26 patients rather than 24 patients in the dalteparin group had experienced confirmed pulmonary embolus, the *P* value would be greater than .05.[8]

When an FI score is 0, this indicates that it is not necessary to move a nonevent to an event in any participant group to remove statistical significance from the study. In this case, the study is very fragile.

A number of online calculators are readily available for FI score calculation.[9,10] The FQ is calculated by dividing the FI score by the total sample size of the trial.[11]

Limitations of the FI Score

The FI is only appropriate for RCTs. The main problem in observational and retrospective studies is the presence of confounders and selection bias, which are prevented in RCTs through randomization and blinding. Advanced statistical tools to enable reliable control over confounders, such as propensity scores and sensitivity analysis, can reduce the bias caused by the lack of randomization in observational and retrospective studies. However, given their considerable limitations, the FI score should be used only in RCTs.

The FI score is also only appropriate for dichotomous outcomes, and it cannot be applied to an outcome that is a continuous variable. Similarly, the FI score may not be appropriate for time-to-event binary outcomes, because it does not account for differences in outcome over time. It can be used in time-to-event binary outcome data in RCTs in which the primary outcome measure is the number of events in each group (as in the PROTECT trial), rather than the timing of the events. However, applying the FI score to time-to-event data in which the numbers of events in both groups are similar but a clear difference in event timing exists may be inappropriate. Use in these cases might result in finding such trials overly fragile.

Table 1. Fragility Index Calculation Example, Based on the Prophylaxis for Thromboembolism in Critical Care Trial (PROTECT) Study[7]

Study Cohort (N = 1873)	Pulmonary Embolism	No Pulmonary Embolism	*P* Value
Original results, No. (%)			
Intervention group	24 (1.3)	1849 (98.7)	.01
Control group	43 (2.3)	1830 (97.7)	
First step of Fragility Index calculation			
Intervention group	25	1848	.04
Control group	43	1830	
Second step of Fragility Index calculation			
Intervention group	26	1847	.05
Control group	43	1830	

Currently, there is no specific FI value that defines an RCT outcome as robust and no FI score cutoff value considered acceptable. Trials with lower FI scores are more fragile (which is usually in association with the smaller number of events, smaller sample size, and resulting lower study power), and trials with a higher FI score are less fragile (which is usually associated with larger number of events, larger sample size, and resulting higher study power). The results of a trial should be viewed with particular skepticism if the number of patients who are lost to follow-up is greater than the FI score, because factoring in the unknown outcomes of these patients could alter the significant result.

Finally, because RCTs are powered to detect the treatment effect for the primary outcome measure, the use of FI scores to assess secondary outcome measures in studies may be limited. Although the FI investigates how sensitive a study's interpretation is compared with the number of observed events, particularly in the context of patient drop out and/or loss to follow-up, it should not be used as a sole measure of the strength of the effect.[12-14]

FI Scores and Clinical Care Guidelines

It has now been recognized that RCTs that provide the foundation for treatment guideline recommendations may not be robust. A review[15] of the 21 RCTs that were used to support treatment recommendations in the 2016 "Chest Guideline and Expert Panel Report on Antithrombotic Therapy for VTE Disease" found a median (interquartile range [IQR]) FI score of 5 (1-9) and a median (IQR) FQ of 0.012 (0.002-0.032), confirming that some RCTs used to support guideline recommendations are fragile. A review of 35 RCTs in the 2017 diabetes treatment guidelines reported the median (IQR) FI score was 16 (4-29), with a median (IQR) FQ score of 0.007 (0.003-0.014), confirming moderate robustness and significant variability.[16]

In contrast to these results, a review of 25 RCTs in heart failure reported a median sample size of 2331 patients. The median (range) FI score was 26 (0-118) for the primary outcome measure in 20 trials. The FI score was 10 or less in only 7 of the RCTs (35%), and in only 4 trials (20%) did the number of patients lost to follow-up exceed the FI score.[17] Importantly, these authors also confirmed a lack of correlation between the treatment effect size for the primary end point and FI score, confirming that the treatment effect size is an unreliable measure of the robustness of RCT findings.

Table 2. Clinical Trauma Trials and Fragility Index Scores

Trial	Patients, No.	Primary Outcome	Primary Outcome Fragility Index Score	Fragility Quotient Score	Secondary Outcome	Secondary Outcome Fragility Index Score
Shakur et al,[18] 2010	20 207	4-wk All-cause mortality	51	0.0025	Vascular occlusive events	NS
Hutchinson et al,[19] 2016	389	Death at 6 mo	24	0.0620	Death at 12 mo	22
Fehlings et al,[20] 2012	313	>2 Grade improvement in American Spinal Injury Association impairment scale grade at 6 mo	13	0.0415		
Eckelt et al,[21] 2006	88	Terminal/lateral shifts	8	0.0909	Pain-free status at 6 mo	3
Goldenstein et al,[22] 2015	181	Effective hemostasis	7	0.0386		
Tanaka et al,[23] 2002	37	Pneumonia	4	0.1081	Return to work at 6 mo	5
Cooper et al,[24] 2011	155	Unfavorable outcomes at 6 mo	3	0.0194	Unfavorable score	3
Andrews et al,[25] 2015	387	Favorable outcomes by Erasmus Guillain-Barré syndrome Outcome Score	3	0.0078	Serious adverse events	9
Stannard et al,[26] 2009	59	Total deep Infections	2	0.0339		
Bickell et al,[27] 1994	598	Hospital survival	1	0.0017	Complications	0
Vidán et al,[28] 2005	319	Mortality	1	0.0031	Developed major complication	8
Marion et al,[29] 1997	82	Favorable Glasgow Outcome Scale-Extended at 12 mo	1	0.0122	Favorable Glasgow Outcome Scale-Extended at 12 mo in patients with Glasgow Coma Scales of 5 to 7	3
Morrison et al,[30] 2011	90	Died within 24 h of admission owing to exsanguination	1	0.0111	30-D mortality	NS
Gonzalez et al,[31] 2016	111	Survival at 28 d	0	NA	Death in 6 h	1
Holcomb et al,[32] 2015	680	Mortality at 24 h and 30 d	NS	NA	Death from exsanguination within 24 h	2
Bhandari et al,[33] 2015	2447	Reoperation rate in groups with irrigation pressure	NS	NA	Reoperation rate in soap vs saline	6
Dutton et al,[34] 2002	110	In-hospital mortality	NS	NA	NA	NA
Smith et al,[35] 2017	103	Mortality at 30 d	NS	NA	Development of intra-abdominal abscess	0
The WOMAN Trial Collaborators,[36] 2017	20 060	Mortality at 42 d	NS	NA	Death owing to bleeding	0
Bulger et al,[37] 2004	46	Pneumonia	NS	NA	NA	NA
Frobell et al,[38] 2010	121	Change in Knee Injury Osteoarthritis Outcome score	Not binary outcome	NA	NA	NA
Young et al,[39] 2014	65	Mean change in base excess between 0 and 24 h	Not binary outcome	NA	NA	NA
Bernard et al,[40] 2010	312	Median Glasgow Outcome Scale-Extended score at 6 mo	Not binary outcome	NA	Favorable Glasgow Outcome Scale-Extended score at 6 mo	1
Marasco et al,[41] 2013	46	Duration of ventilation and intensive care unit length of stay	Not binary outcome	NA	Received noninvasive ventilation postextubation	0
Meyer et al,[42] 1997	39	Hospital length of stay	Not binary outcome	NA	NA	NA

Abbreviation: NA, not applicable; NS, not significant.

The FI Score in Trauma Trials

No studies, to our knowledge, have yet examined FI scores in trials on trauma topics. This is important, because the results of RCTs on trauma are used to develop recommendations for best practices and guidelines for trauma care. None of the current national and international trauma guidelines have incorporated assessment of FI scores into the assessment of RCTs included in the final guideline recommendations.

We systematically reviewed trauma RCTs eligible for the calculation of FI score for either the primary or secondary outcome measures. We identified 25 eligible RCTs, of which only 14 (56%) had statistically significant results for the primary outcome measure; these 14 had a median (IQR) sample size of 168 (89-370) patients (**Table 2**). The median (IQR) FI score was 3 (1-8), which means that adding 3 events to the opposite treatment arm eliminated statistical significance. The median FQ score was 0.016 (0.0043-0.0408).

The Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage (CRASH-2) trial[43] (NCT00375258; Table 2) had an FI score of 51 for the primary outcome measure of 4-week all-cause mortality, which was the highest of all trauma trials reviewed, but 84 patients were lost to follow-up. A number of patients lost to follow-up higher than the FI score detracts from usefulness of translating the findings of this study into clinical practice. This issue,

along with other trial weaknesses, have limited the empirical use of tranexamic acid in US trauma centers.[43]

Interestingly, results from the Pragmatic, Randomized Optimal Platelet and Plasma Ratios (PROPPR) trial[32] (NCT01545232; Table 2) confirmed that among 680 patients with severe trauma and major bleeding, there was no significant difference in 24-hour or 30-day mortality in patients who received a blood product transfusion of plasma, platelets, and red blood cells in a ratio of 1:1:1 compared with those who received transfusions with a ratio of 1:1:2. The study was significantly underpowered for the observed difference in mortality found; 2968 patients would need to be enrolled in the study to detect a difference of 4.2% in 24-hour mortality with 90% power. However, a ratio of 1:1:1 resulted in reduced incidence of death from exsanguination within the first 24 hours (9.2% vs 14.6%; difference, −5.4% [95% CI, −10.4% to −0.5%]; *P* = .03) as a secondary outcome measure, with a calculated FI score of 2. Although the PROPPR study results are deemed fragile, recommendations for 1:1:1 transfusion strategies are recommended in national trauma guidelines.[44]

Similarly, a single-institution RCT (n = 111) comparing hemostatic resuscitation with viscoelastic testing and hemostatic resuscitation with conventional coagulation assays (Table 2) documented significantly decreased 28-day mortality (the primary outcome measure) in the group that received thromboelastography (11 vs 20 deaths; *P* = .049), but the FI score was 0. For the secondary outcome measure of death in 6 hours, the group who received thromboelastography had decreased mortality (4 vs 12 deaths; *P* = .03), but the FI score was 2. No difference in hemorrhagic deaths (5 vs 11; *P* = .11) was identified. In "Resources for Optimal Care of the Injured Patient,"[45] the Committee on Trauma of the American College of Surgeons recommends that thromboelastography be available at level I and II trauma centers.

We have confirmed that in trauma RCTs reporting statistically significant results, the sample sizes were small in most studies, and the findings often were dependent on a very small number of events. Based on these findings, there is a considerable opportunity in the future to design trauma RCTs with larger sample and effect sizes to improve the robustness of trauma trials.

The FI Score in Medical and Surgical Trials

A review[46] of 399 RCTs in high-impact medical journals (eg, *JAMA*, *the New England Journal of Medicine*, *the Lancet*) reported that 53% of the total had a *P* value less than .01, but the median (range) FI score was 8 (0-109). Importantly, 25% of the RCTs had an FI score of 3 or less, and in 53%, the FI score was less than the number of patients lost to follow-up. Furthermore, 10% of the trials had an FI score of 0. These data confirmed that the statistically significant results of many RCTs hinge on small numbers of events.

A recent systematic review[47] examined the FI score of 56 RCTs on critical care topics that reported a statistically significant effect on mortality. The median (IQR) FI score was 2 (1-3.5). A total of 24 trials (43%) had an FI score of 0 or 1, with 11 trials (20%) with an FI score of 0 and 13 trials (23%) with an FI score of 1. Seven trials (13%) reported loss to follow-up greater than their FI score.[47] In a study of 127 RCTs in nephrology, 20 (15.7%) had an FI score of 0, and the median (range) FI score was 3 (1-166), indicating that, in half of the RCTs, 3 events in the treatment group would render the trial results nonsignificant.[48]

A systematic survey of RCTs on spine surgery topics identified 40 RCTs with a median (IQR) FI score of 2 (1-3), and the FI score was 3 or less in 75% of the trials. The FI score was less than or equal to the number of patients lost to follow-up in 26 of the trials (65%).[49] Similar findings were identified in a review of 48 RCTs on sports surgery topics, with a median (IQR) FI score of 2 (1-2.8), confirming that most significant RCT findings in sports medicine and arthroscopic surgery are fragile.[50] Similar findings were reported for RCTS on head and neck surgery (27 trials; median [IQR] FI score, 1 [0-2.5]),[51] urological surgery (41 trials; median [IQR] FI score, 3 [1-4.5]),[52] intracranial hemorrhage (98 trials; median [IQR] FI score, 2 [1-4]), and ophthalmology (156 trials; median [IQR] FI score, 2 [0-48]).[53]

Similarly, a systematic review of 139 anesthesiology RCTs[54] reported a median (IQR) FI score of 4 (2-17) for 35 RCTs published in general medicine journals vs a median (IQR) score of 3 (2-7) for 104 RCTs published in anesthesiology journals. Patients lost to follow-up exceeded the FI score in 56 of 139 trials (40.3%), including 14 of 35 (40%) in general medicine journals and 28 of 104 (26.9%) in anesthesiology journals. These results confirm that most anesthesiology RCTs are fragile, with particular concern that the primary outcome results may have changed based on the results of the patients lost to follow-up.[54]

How to Use the FI Score in Clinical Care

Many trauma and surgical RCTs have inadequate sample sizes to accrue an adequate number of events to assess the treatment effect. Surgeons should understand that the sole use of threshold *P* values, hazard ratios, and 95% CIs to assess statistical significance in RCTs is inadequate to assess the robustness of RCTs. The American Statistical Association has issued a statement stressing the principle that "a *P* value, or statistical significance, does not measure the size of an effect or the importance of a result."[55] The FI provides an easy additional metric for clinicians to use in assessing the treatment effect on patient outcomes.

Surgeons should particularly exercise caution when considering a potential change in clinical practice based on RCTs with a low FI score. If the number of patients lost to follow-up is greater than the FI score, the RCT findings should be viewed with considerable caution.

Conclusions

The FI is an important aid to the clinician's interpretation of clinical trial results. An FI score measures the robustness (or fragility of the results of a clinical trial and is important for optimal interpretation of all randomized clinical trial results. The FI score complements the *P* value and helps identify less robust results.[46]

The results presented confirm that most RCTs in surgery and medicine are fragile, in contrast with most RCTs in cardiac disease and heart failure, which are robust. The provision of high-quality, evidence-based clinical care in surgery for optimal patient outcomes requires a foundation of robust clinical research evidence, and knowledge of the FI score will assist in future surgical RCT design.

We strongly recommend the routine reporting of FI scores for all trauma and surgical RCTs to assist in appropriate and optimal decision-making in the care of patients who have experienced trauma and/or need surgery. We also recommend the routine inclusion of FI scores in the development of clinical guidelines to assist clinicians in ascertaining whether guideline recommendations are robust. Surgeons should be aware to particularly exercise caution when considering a potential change in clinical practice based on RCTs with a low FI score.

ARTICLE INFORMATION

Accepted for Publication: October 5, 2018.

Published Online: November 7, 2018. doi:10.1001/jamasurg.2018.4318

Author Contributions: Drs Tignanelli and Napolitano had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: All authors.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: All authors.

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Sterne JA, Davey Smith G. Sifting the evidence—what's wrong with significance tests? *BMJ*. 2001;322(7280):226-231. doi:10.1136/bmj.322.7280.226

2. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. doi:10.1371/journal.pmed.0020124

3. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004;96(6):434-442. doi:10.1093/jnci/djh075

4. Risch NJ. Searching for genetic determinants in the new millennium. *Nature*. 2000;405(6788):847-856. doi:10.1038/35015718

5. Walter SD. Statistical significance and fragility criteria for assessing a difference of two proportions. *J Clin Epidemiol*. 1991;44(12):1373-1378. doi:10.1016/0895-4356(91)90098-T

6. Feinstein AR. The unit fragility index: an additional appraisal of "statistical significance" for a contrast of two proportions. *J Clin Epidemiol*. 1990;43(2):201-209. doi:10.1016/0895-4356(90)90186-S

7. Cook D, Meade M, Guyatt G, et al; PROTECT Investigators for the Canadian Critical Care Trials Group and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Dalteparin versus unfractionated heparin in critically ill patients. *N Engl J Med*. 2011;364(14):1305-1314. doi:10.1056/NEJMoa1014475

8. Cook D, Meade M, Guyatt G, et al; PROTECT Investigators for the Canadian Critical Care Trials Group and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Dalteparin versus unfractionated heparin in critically ill patients. *N Engl J Med*. 2011;364(14):1305-1314. doi:10.1056/NEJMoa1014475

9. ClinCalc LLC. Fragility index calculator. http://clincalc.com/Stats/FragilityIndex.aspx. Published 2018. Accessed September 25, 2018.

10. Spooner F, Ridgeon EE. Fragility index. http://www.fragilityindex.com/. Published 2016. Accessed September 25, 2018.

11. Ahmed W, Fowler RA, McCredie VA. Does sample size matter when interpreting the fragility index? *Crit Care Med*. 2016;44(11):e1142-e1143. doi:10.1097/CCM.0000000000001976

12. Carter RE, McKie PM, Storlie CB. The Fragility Index: a *P* value in sheep's clothing? *Eur Heart J*. 2017;38(5):346-348.

13. Walsh M, Devereaux PJ, Sackett DL. Clinician trialist rounds: 28. When RCT participants are lost to follow-up, part 1: why even a few can matter. *Clin Trials*. 2015;12(5):537-539. doi:10.1177/1740774515597702

14. Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ*. 2012;344:e2809. doi:10.1136/bmj.e2809

15. Edwards E, Wayant C, Besas J, Chronister J, Vassar M. How fragile are clinical trial outcomes that support the CHEST clinical practice guidelines for VTE? *Chest*. 2018;154(3):512-520. doi:10.1016/j.chest.2018.01.031

16. Chase Kruse B, Matt Vassar B. Unbreakable? An analysis of the fragility of randomized trials that support diabetes treatment guidelines. *Diabetes Res Clin Pract*. 2017;134:91-105. doi:10.1016/j.diabres.2017.10.007

17. Docherty KF, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. How robust are clinical trials in heart failure? *Eur Heart J*. 2017;38(5):338-345.

18. Shakur H, Roberts I, Bautista R, et al; CRASH-2 trial collaborators. Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial. *Lancet*. 2010;376(9734):23-32. doi:10.1016/S0140-6736(10)60835-5

19. Hutchinson PJ, Kolias AG, Timofeev IS, et al; RESCUEicp Trial Collaborators. Trial of decompressive craniectomy for traumatic intracranial hypertension. *N Engl J Med*. 2016;375(12):1119-1130. doi:10.1056/NEJMoa1605215

20. Fehlings MG, Vaccaro A, Wilson JR, et al. Early versus delayed decompression for traumatic cervical spinal cord injury: results of the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS). *PLoS One*. 2012;7(2):e32037. doi:10.1371/journal.pone.0032037

21. Eckelt U, Schneider M, Erasmus F, et al. Open versus closed treatment of fractures of the mandibular condylar process-a prospective randomized multi-centre study. *J Craniomaxillofac Surg*. 2006;34(5):306-314. doi:10.1016/j.jcms.2006.03.003

22. Goldstein JN, Refaai MA, Milling TJ Jr, et al. Four-factor prothrombin complex concentrate versus plasma for rapid vitamin K antagonist reversal in patients needing urgent surgical or invasive interventions: a phase 3b, open-label, non-inferiority, randomised trial. *Lancet*. 2015;385(9982):2077-2087. doi:10.1016/S0140-6736(14)61685-8

23. Tanaka H, Yukioka T, Yamaguti Y, et al. Surgical stabilization of internal pneumatic stabilization? a prospective randomized study of management of severe flail chest patients. *J Trauma*. 2002;52(4):727-732.

24. Cooper DJ, Rosenfeld JV, Murray L, et al; DECRA Trial Investigators; Australian and New Zealand Intensive Care Society Clinical Trials Group. Decompressive craniectomy in diffuse traumatic brain injury. *N Engl J Med*. 2011;364(16):1493-1502. doi:10.1056/NEJMoa1102077

25. Andrews PJ, Sinclair HL, Rodriguez A, et al; Eurotherm3235 Trial Collaborators. Hypothermia for intracranial hypertension after traumatic brain injury. *N Engl J Med*. 2015;373(25):2403-2412. doi:10.1056/NEJMoa1507581

26. Stannard JP, Volgas DA, Stewart R, McGwin G Jr, Alonso JE. Negative pressure wound therapy after severe open fractures: a prospective randomized study. *J Orthop Trauma*. 2009;23(8):552-557. doi:10.1097/BOT.0b013e3181a2e2b6

27. Bickell WH, Wall MJ Jr, Pepe PE, et al. Immediate versus delayed fluid resuscitation for hypotensive patients with penetrating torso injuries. *N Engl J Med*. 1994;331(17):1105-1109. doi:10.1056/NEJM199410273311701

28. Vidán M, Serra JA, Moreno C, Riquelme G, Ortiz J. Efficacy of a comprehensive geriatric intervention in older patients hospitalized for hip fracture: a randomized, controlled trial. *J Am Geriatr Soc*. 2005;53(9):1476-1482. doi:10.1111/j.1532-5415.2005.53466.x

29. Marion DW, Penrod LE, Kelsey SF, et al. Treatment of traumatic brain injury with moderate hypothermia. *N Engl J Med*. 1997;336(8):540-546. doi:10.1056/NEJM199702203360803

30. Morrison CA, Carrick MM, Norman MA, et al. Hypotensive resuscitation strategy reduces transfusion requirements and severe postoperative coagulopathy in trauma patients with hemorrhagic shock: preliminary results of a randomized controlled trial. *J Trauma*. 2011;70(3):652-663. doi:10.1097/TA.0b013e31820e77ea

31. Gonzalez E, Moore EE, Moore HB, et al. Goal-directed hemostatic resuscitation of trauma-induced coagulopathy: a pragmatic randomized clinical trial comparing a viscoelastic assay to conventional coagulation assays. *Ann Surg*. 2016;263(6):1051-1059. doi:10.1097/SLA.0000000000001608

32. Holcomb JB, Tilley BC, Baraniuk S, et al; PROPPR Study Group. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA*. 2015;313(5):471-482. doi:10.1001/jama.2015.12

33. Bhandari M, Jeray KJ, Petrisor BA, et al; FLOW Investigators. A trial of wound irrigation in the initial management of open fracture wounds. *N Engl J Med*. 2015;373(27):2629-2641. doi:10.1056/NEJMoa1508502

34. Dutton RP, Mackenzie CF, Scalea TM. Hypotensive resuscitation during active

hemorrhage: impact on in-hospital mortality. *J Trauma*. 2002;52(6):1141-1146. doi:10.1097/00005373-200206000-00020

35. Smith JW, Matheson PJ, Franklin GA, Harbrecht BG, Richardson JD, Garrison RN. Randomized controlled trial evaluating the efficacy of peritoneal resuscitation in the management of trauma patients undergoing damage control surgery. *J Am Coll Surg*. 2017;224(4):396-404. doi:10.1016/j.jamcollsurg.2016.12.047

36. WOMAN Trial Collaborators. Effect of early tranexamic acid administration on mortality, hysterectomy, and other morbidities in women with post-partum haemorrhage (WOMAN): an international, randomised, double-blind, placebo-controlled trial. *Lancet*. 2017;389(10084):2105-2116. doi:10.1016/S0140-6736(17)30638-4

37. Bulger EM, Edwards T, Klotz P, Jurkovich GJ. Epidural analgesia improves outcome after multiple rib fractures. *Surgery*. 2004;136(2):426-430. doi:10.1016/j.surg.2004.05.019

38. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS. A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med*. 2010;363(4):331-342. doi:10.1056/NEJMoa0907797

39. Young JB, Utter GH, Schermer CR, et al. Saline versus Plasma-Lyte A in initial resuscitation of trauma patients: a randomized trial. *Ann Surg*. 2014;259(2):255-262. doi:10.1097/SLA.0b013e318295feba

40. Bernard SA, Nguyen V, Cameron P, et al. Prehospital rapid sequence intubation improves functional outcome for patients with severe traumatic brain injury: a randomized controlled trial. *Ann Surg*. 2010;252(6):959-965. doi:10.1097/SLA.0b013e3181efc15f

41. Marasco SF, Davies AR, Cooper J, et al. Prospective randomized controlled trial of operative rib fixation in traumatic flail chest. *J Am Coll Surg*. 2013;216(5):924-932. doi:10.1016/j.jamcollsurg.2012.12.024

42. Meyer DM, Jessen ME, Wait MA, Estrera AS. Early evacuation of traumatic retained hemothoraces using thoracoscopy: a prospective, randomized trial. *Ann Thorac Surg*. 1997;64(5):1396-1400. doi:10.1016/S0003-4975(97)00899-0

43. Napolitano LM, Cohen MJ, Cotton BA, Schreiber MA, Moore EE. Tranexamic acid in trauma: how should we use it? *J Trauma Acute Care Surg*. 2013;74(6):1575-1586. doi:10.1097/TA.0b013e318292cc54

44. Cannon JW, Khan MA, Raja AS, et al. Damage control resuscitation in patients with severe traumatic hemorrhage: a practice management guideline from the Eastern Association for the Surgery of Trauma. *J Trauma Acute Care Surg*. 2017;82(3):605-617. doi:10.1097/TA.0000000000001333

45. Committee on Trauma American College of Surgeons. Resources for optimal care of the injured patient. https://www.facs.org/~/media/files/quality%20programs/trauma/vrc%20resources/resources%20for%20optimal%20care.ashx.Published 2014. Accessed September 25, 2018.

46. 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

47. 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

48. Shochet LR, Kerr PG, Polkinghorne KR. The fragility of significant results underscores the need of larger randomized controlled trials in nephrology. *Kidney Int*. 2017;92(6):1469-1475. doi:10.1016/j.kint.2017.05.011

49. 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(10):2188-2197. doi:10.1016/j.spinee.2015.06.004

50. 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(9):2164-2170. doi:10.1177/0363546516674469

51. Noel CW, McMullen C, Yao C, et al. The fragility of statistically significant findings from randomized trials in head and neck surgery [published online April 23, 2018]. *Laryngoscope*. doi:10.1002/lary.27183

52. 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

53. Shen C, Shamsudeen I, Farrokhyar F, Sabri K. Fragility of results in ophthalmology randomized controlled trials: a systematic review. *Ophthalmology*. 2018;125(5):642-648. doi:10.1016/j.ophtha.2017.11.015

54. 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(5):935-941. doi:10.1016/j.bja.2018.01.012

55. Wasserstein RL, Lazar NA. The American Statistical Association (ASA)'s statement on *P* values: context, process, and purpose. *Am Stat*. 2016;70:129-133. doi:10.1080/00031305.2016.1154108