How Robust are the Evidences that Formulate Surviving Sepsis Guidelines? An Analysis of Fragility and Reverse Fragility of Randomized Controlled Trials that were Referred in these Guidelines

Nang S Choupoo1, Saurabh K Das2, Priyam Saikia3, Samarjit Dey4, Sumit Ray5

ABSTRACT

Objectives: “Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016” provides guidelines in regard to prompt management and resuscitation of sepsis or septic shock. The study is aimed to assess the robustness of randomized controlled trials (RCTs) that formulate these guidelines in terms of fragility index and reverse fragility index.

Method: RCTs that contributed to these guidelines having parallel two-group design, 1:1 allocation ratio, and at least one dichotomous outcome were included in the study. The median fragility index was calculated for RCTs with significant statistical outcomes, whereas the median reverse fragility index was calculated for RCTs with nonsignificant statistical results.

Results: Hundred RCTs that met the inclusion criteria were analyzed. The median fragility index was 5.5 (95% confidence interval (CI) 1–30) and median reverse fragility index was 13 (95% CI 12.07–16.8) at a p value of 0.05. The median reverse fragility index was 16 (95% CI 10–26) at a p value of 0.01. Most of the RCTs included in this analysis were of good quality, having a median Jadad score of 6.

Conclusion: This analysis found that the surviving sepsis guidelines were based on highly robust RCTs with statistically insignificant results and on some moderately robust RCTs with statistically significant results. RCTs with statistically insignificant results were more robust than RCTs with statistically significant results in regard to these guidelines.

Keywords: Fragility index, Revised fragility index, Surviving sepsis guidelines.

Highlights: The study assessed the robustness of randomized controlled trials (RCTs) that were used to formulate surviving sepsis guidelines. Most RCTs showed statistically nonsignificant results. RCTs with statistically significant results were moderately fragile whereas RCTs with nonsignificant results were more robust.

Indian Journal of Critical Care Medicine (2021): 10.5005/jp-journals-10071-23895

Introduction

The probability values, more popularly known as p values, are widely used to quantify the statistical significance of observed results. The practice of significance testing originated from the concept and practice of the renowned statistician, R.A. Fisher, in the third decade of the 20th century.1 However, p values have been frequently subjected to criticism due to its potential misinterpretation. When a p value was introduced, it was not supposed to be used as a definitive test but was a casual way to determine whether the evidence was significant in an old-fashioned way. It is often assumed that a lower p value indicates a more statistically significant result. Many erroneously regard statistical significance as having clinical significance. This is an oversimplification and may result in overemphasis on the clinical importance of the study. A large study could have the same p value as a very small study. While both are regarded as “statistically significant,” the p value does not provide any indication that there is a clear distinction between these studies, leading one to conclude that the likelihood of a true effect is the same. Another important fallacy is that only one event can make a significant result nonsignificant and vice versa. The former is typically interpreted as indicating a more important treatment effect, although there being minimum absolute difference between the two types of result.2,3

Therefore, to decrease the absolute reliance on p value, various measures have been postulated, and they are lowering p value threshold, using alternative approaches like effect size and confidence interval, Bayes factor, Akaike information criterion, incorporation of fragility index (FI), etc.4–6 The concept of fragility was introduced by Feinstein in the epidemiology literature.7

1Department of Anesthesia, Atal Bihari Vajpayee Medical Institute and Dr RML Hospital, Delhi, India
2Department of Critical Care Medicine, Artemis Hospital, Gurugram, Haryana, India
3Department of Anaesthesiology and Critical Care, Gauhati Medical College and Hospital, Guwahati, Assam, India
4Department of Anaesthesia and Critical Care, AIIMS, Raipur, Chhattisgarh, India
5Department of Critical Care, Holy Family Hospital, Delhi, India

Corresponding Author: Saurabh K Das, Department of Critical Care Medicine, Artemis Hospital, Gurugram, Haryana, India, Phone: +91 8587889525, e-mail: dassk1729@gmail.com

How to cite this article: Choupoo NS, Das SK, Saikia P, Dey S, Ray S. How Robust are the Evidences that Formulate Surviving Sepsis Guidelines? An Analysis of Fragility and Reverse Fragility of Randomized Controlled Trials that were Referred in these Guidelines. Indian J Crit Care Med 2021;25(7):773-779.

Source of support: Nil
Conflict of interest: None
This implies the minimum number of patients whose status would have to be changed from a "nonevent" to an "event" in order to turn a statistically significant result into a nonsignificant result. If lesser numbers are required to change the statistical significance of the study, it is regarded to be the lack of robustness of a trial result. FI is exclusively applied to trials that reach traditional statistical significance. To check the robustness of a statistically nonsignificant trial, reverse fragility index (RFI) has been used. RFI provides a measure of robustness in the neutrality of results when assessed from a clinical perspective.

"Surviving Sepsis Campaign: International Guidelines for management of Sepsis and Septic Shock: 2016" provided 93 statements on early management and resuscitation of patients with sepsis or septic shock. These guidelines are a careful synthesis of available randomized controlled trials (RCTs), systematic review and meta-analysis, and case-control studies that encompass a wide range of management strategies including early resuscitation, goal-directed therapy, antibiotic therapy, fluid therapy, vasoactive medications, corticosteroids, immunoglobulins, blood purifications, anticoagulants, mechanical ventilation, sedation analgesia, glucose control, renal replacement therapy, etc. The purpose of this study is to apply FI and RFI analysis to the latest surviving sepsis guidelines (SSG) and to assess the fragility of RCTs, reporting dichotomous outcome parameters.

**Materials and Methods**

**Data Search**
Recent Surviving Sepsis Campaign guidelines published in the year 2016 were reviewed. Two independent investigators (SKD and NSG) screened all the RCTs referenced in guidelines and assessed them for inclusion. Any disagreement was resolved by consensus with a third author (PS).

**Eligibility Criteria**
- RCTs with parallel two-group design
- 1:1 allocation ratio
- At least one dichotomous outcome was included in the study.

Letters, editorials, systematic reviews or meta-analyses, opinions, observational studies, economic or cost-effective analyses of RCTs, cohort nonrandomized studies, and quasi-randomized trials were excluded.

**Data Collection**
A prespecified data collection form was used to extract the following data from all RCTs: studied intervention, authors, binary outcomes, sample sizes, number of patients with events, and number of patients without events. We prioritized the primary outcomes for the analysis; however, when analyzable data were not available, secondary dichotomous outcomes related to mortality were included.

**Quality Assessment**
Quality assessment of included studies was done by one investigator (PS) using "modified Jadad scale." A questionnaire based eight questions was used to assess randomization, blinding withdrawal or dropouts, description of inclusion/exclusion criteria, assessment of adverse effects, and description of the statistical plan. A score of 1 to 8 was given to each study where 8 denotes maximum robustness whereas 1 denotes least.

**Outcomes**
The outcomes were FI and RFI at p values of 0.05 and 0.01, fragility quotient (FQ) and reverse fragility quotient (RFQ).

**Statistical Analysis**
For each included outcome from RCTs, a two-by-two contingency table was created. FI was calculated according to the method described by Walsh et al. The number of events was added to a group with a smaller number of events while subtracting nonevents from the same group to keep the total number of participants constant. Events were added iteratively and calculations were done with a Fisher's exact test for each addition until the calculated p value became just more than 0.05. RFI was calculated according to the method described in a recent publication. The RFI was calculated by subtracting events from the group with a lower number of events while simultaneously adding nonevents to the same group to keep the number of participants constant until the Fisher's exact test two-sided p value became less than 0.05. A similar method was used to calculate RFI at a p value of 0.01.

FI or RFI is an absolute measure of stability, irrespective of trial size. We analyzed FQ and RFQ as a relative measure of fragility. This was calculated by dividing the FI or RFI by its respective sample size.

Subgroup analysis was done to analyses FI and RFI of studies testing similar domains of sepsis management, e.g. studies dealt with mechanical ventilation.

FI was calculated using the online FI calculator www.clinicalcalc.com. To calculate a Fisher's exact test two-sided p value, the online calculator https://www.graphpad.com/quickcalcs was used.

**Result**
After screening 655 references of surviving sepsis guidelines 2016 (SSG2016), a total of 201 RCTs were identified. Of these, 100 RCTs were included in the final analysis. Among the included RCTs, 22 had dichotomous statistically significant outcome measures and 78 studies reported statistically insignificant dichotomous outcome measures (Fig. 1). Median sample size of RCTs with significant result was 286 (95% confidence interval (CI) 32–6,104). The median sample size of RCTs with statistically insignificant results was 520 (95% CI 31–6,997) (Tables 1 and 2).

Median FI was 5.5 (95% CI 1.1–30) and median RFI was 13 (95% CI 12.07–16.8) at a p value of 0.05.

Median FQ was 0.01 (95% CI 0.01–0.02) and median RFQ was 0.02 (95% CI 0.02–0.04).

Median RFI was 16 (95% CI 10–26) at a p value of 0.01.

**Quality Assessment**
Most of the RCTs included in this analysis were of good quality. The median Jadad score of RCTs with significant results was 6 (95% CI 5–8) and the median Jadad score of RCTs with nonsignificant results was also 6 (95% CI 4–8).

**Subgroup Analysis**
RCTs that are included in this analysis were grouped according to the domains they dealt with (Table 3). Three most commonly studied subjects that were analyzed by the RCTs were mechanical ventilation, nutrition, and goal-directed therapy. Fifteen studies were done on various ventilator strategies; ECMO and other supportive measures had a median FI and RFI of 4 and 12,
Robustness of Surviving Sepsis Guidelines

Table 1: Characteristics of included studies with statistically significant results

| Studies          | Intervention                                         | Sample size | Fragility index | Fragility quotient | Jadad score |
|------------------|------------------------------------------------------|-------------|-----------------|--------------------|-------------|
| Rivers E         | EGDT                                                 | 263         | 4               | 0.01               | 7.5         |
| Bernard GR       | Recombinant human protein C                          | 1,690       | 15              | 0.008              | 8           |
| de Jong E        | Procalcitonin-guided antibiotic therapy              | 1,546       | 9               | 0.005              | 6           |
| Martin C         | Dopamine vs norepinephrine                          | 32          | 5               | 0.15               | 5           |
| Corwin HL        | Recombinant erythropoietin                           | 1,302       | 30              | 0.20               | 8           |
| Bollaert PE      | Hydrocortisone                                       | 41          | 7               | 0.17               | —           |
| Amato MB         | Protective ventilation                               | 53          | 1               | 0.01               | 6           |
| Brower RG        | Low tidal volume                                     | 861         | 12              | 0.01               | 5           |
| Villar J         | High PEEP, low tidal volume                          | 103         | 1               | 0.009              | 5           |
| Guérin C         | Prone position 14                                    | 466         | 20              | 0.04               | 6           |
| Peek GJ          | ECMO                                                 | 180         | 2               | 0.01               | 6           |
| Ferguson ND      | HFOV                                                 | 548         | 10              | 0.01               | 6           |
| Ferrer M         | NIV                                                  | 105         | 4               | 0.03               | 5           |
| Gao Smith F      | Intravenous β2 agonist in ARDS                       | 326         | 2               | 0.006              | 8           |
| Futier E         | Intraoperative low tidal volume                      | 400         | 17              | 0.04               | 8           |
| Drakulovic MB    | Supine body position                                | 86          | 3               | 0.03               | 5           |
| Schweickert WD   | Early physical and occupational therapy             | 104         | 3               | 0.02               | 6           |
| van den Berghe G | Intensive insulin therapy                            | 1,548       | 7               | 0.004              | 6           |
| Finfer S         | Intensive insulin therapy                            | 6,104       | 9               | 0.001              | 6           |
| Fuentes-Orozco C | L-alanyl-L-glutamine                                  | 33          | 3               | 0.09               | 8           |
| Detering KM      | Advance care planning on end-of-life care           | 309         | 6               | 0.01               | 5           |
| Aguado JM        | Galactomannan and PCR-based DNA detection of aspergillus | 203     | 1               | 0.004              | 6           |

EGDT, early goal-directed therapy; ECMO, extracorporeal membrane oxygenator; HFOV, high-frequency oscillating ventilation; NIV, noninvasive ventilation

respectively. Thirteen studies on nutrition were analyzed; of which 12 studies showed nonsignificant results having a median RFI of 7.5. Eight studies were done on the efficacy of goal-directed therapy; except one all RCTs had nonsignificant results with a median RFI of 6. Subgroup analysis also revealed that studies with insignificant results were more robust than those with significant results.
Table 2: Characteristics of included studies with nonsignificant statistical results

| Author          | Intervention                                                | Sample size | Reverse Fl at p < 0.5 | Reverse Fl at p < 0.01 | Fragility quotient | Jadad score |
|-----------------|-------------------------------------------------------------|-------------|-----------------------|------------------------|--------------------|-------------|
| Peake SL        | Goal-directed resuscitation                                | 1,591       | 28                    | 35                     | 0.01               | 6           |
| Yealy DM        | EGDT                                                        | 895         | 14                    | 20                     | 0.01               | 6           |
| Mouncey PR      | EGDT                                                        | 1,260       | 29                    | 36                     | 0.02               | 6           |
| Hayes MA        | Elevation of oxygen delivery by dobutamine                 | 100         | 1                     | 3                      | 0.005              | 6           |
| Jansen TC       | Lactate-guided resuscitation                               | 348         | 2                     | 7                      | 0.005              | 6           |
| Jones AE        | Lactate vs ScvO\textsubscript{2}-guided resuscitation       | 300         | 6                     | 8                      | 0.02               | 6           |
| Lyu X           | Lactate clearance                                          | 100         | 6                     | 8                      | 0.06               | —           |
| Brunkhorst FM   | Moxifloxacin and meropenem vs meropenem                     | 600         | *13,12                | 18,19                  | 0.02,0.02          | 6           |
| Chastre J       | Eight vs 15 days of antibiotic therapy                     | 401         | 12                    | 15                     | 0.03               | 8           |
| Sawyer RG       | Short-course antimicrobial therapy                          | 517         | 17                    | 23                     | 0.03               | 6           |
| Dunbar LM       | Levofoxacin 750 mg vs 500 mg                                | 528         | 18                    | 25                     | 0.03               | 8           |
| Hepburn MJ      | Short-course antimicrobial therapy                          | 87          | 7                     | 14                     | 0.08               | 8           |
| Rattan R        | Antibiotic duration                                        | 112         | 7                     | 8                      | 0.06               | 6           |
| Caironi P       | Albumin vs crystalloid                                      | 1,818       | 36                    | 45                     | 0.02               | 6           |
| Russell JA      | Vasopressin norepinephrine                                 | 781         | 12                    | 18                     | 0.01               | 8           |
| Gordon AC       | Vasopressin norepinephrine                                 | 408         | 19                    | 24                     | 0.04               | 8           |
| De Backer D     | Dopamine vs norepinephrine                                 | 1,679       | 21                    | 35                     | 0.004              | 8           |
| Annane D        | Epinephrine vs norepinephrine plus dobutamine              | 330         | 12                    | 16                     | 0.03               | 8           |
| Gordon AC       | Levosimendan                                               | 516         | 10                    | 14                     | 0.02               | 8           |
| Briegel J       | Hydrocortisone                                             | 40          | 5                     | 8                      | 0.1                | —           |
| Sprung CL       | Hydrocortisone                                             | 233         | 11                    | 13                     | 0.04               | 8           |
| Annane D        | Hydrocortisone and fludrocortisone                         | 299         | 10                    | 12                     | 0.03               | 8           |
| Huh JW          | Corticosteroids                                            | 130         | 11                    | 12                     | 0.07               | 6           |
| Keh D           | Corticosteroids                                            | 340         | 13                    | 15                     | 0.03               | 8           |
| Holst LB        | Transfusion threshold                                      | 998         | 22                    | 30                     | 0.02               | 7.5         |
| Zumberg MS      | Platelet transfusion                                       | 159         | 6                     | 8                      | 0.04               | 5           |
| Stanworth SJ    | Platelet transfusion                                       | 600         | 2                     | 8                      | 0.02               | 6           |
| Werdan K        | Immunoglobulin G                                           | 624         | 18                    | 23                     | 0.03               | 7           |
| Payen DM        | Polymyxin hemoperfusion                                    | 243         | 10                    | 12                     | 0.04               | 6           |
| Livigni S       | Plasma filtration adsorption                               | 184         | 12                    | 15                     | 0.07               | 6           |
| Warren BL       | Antithrombin III                                           | 2,314       | 46                    | 58                     | 0.02               | 8           |
| Vincent JL      | Thrombomodulin                                             | 741         | 7                     | 12                     | 0.02               | 8           |
| Ranieri VM      | Drotrecogin alfa                                           | 1,680       | 17                    | 25                     | 0.01               | 8           |
| Papazian L      | Cisatracurium infusion in ARDS                              | 339         | 4                     | 6                      | 0.02               | 8           |
| Brochard L      | Reduction of tidal volume                                  | 116         | 7                     | 9                      | 0.06               | 6           |
| Brower RG       | Lower PEEP vs higher PEEP                                   | 549         | 13                    | 18                     | 0.02               | 5           |
| Mercat A        | PEEP                                                       | 767         | 13                    | 17                     | 0.02               | 6           |
| Guerin C        | Prone position                                             | 791         | 22                    | 28                     | 0.03               | 6           |
| Young D         | HFOV                                                       | 795         | 25                    | 30                     | 0.03               | 6           |
| Meade MO        | Low TV, recruitment maneuvers, and high PEEP               | 983         | 11                    | 18                     | 0.01               | 6           |
| Antonelli M     | NIV                                                        | 64          | 6                     | 9                      | 0.03               | 6           |
| Frat JP         | HFNC                                                       | 200         | 6                     | 9                      | 0.03               | 6           |
| Wiedemann HP    | Conservative vs liberal fluid management                   | 1,000       | 14                    | 20                     | 0.01               | 6           |

(Contd...)
science in order to assess the robustness of findings amid concern over the reproducibility of research. A retrospective analysis calculated a median Fl of 56 RCTs in critical care medicine reporting mortality. The median Fl was 2 with an interquartile range (IQR) of 1 to 35. Similar to our study, several clinical guidelines were subjected to Fl analysis. An analysis of 32 RCTs included in the American College of Gastroenterology Guidelines of Crohn’s disease reported a median Fl of 3. An analysis of 21 RCTs that were used to support treatment recommendations in the 2016 Chest

**Table 2: (Contd...)**

| Author                | Intervention                                | Sample size | Reverse Fl at p <0.5 | Reverse Fl at p <0.01 | Fragility quotient | Jadad score |
|-----------------------|---------------------------------------------|-------------|----------------------|-----------------------|--------------------|-------------|
| Wheeler AP            | PAC vs CVC                                  | 1,001       | 21                   | 27                    | 0.02               | —           |
| Richard C             | Pulmonary artery catheter                   | 676         | 21                   | 26                    | 0.02               | 6           |
| Harvey S              | Pulmonary artery catheter                   | 1,041       | 17                   | 22                    | 0.02               | 6           |
| Rhodes A              | Pulmonary artery catheter                   | 201         | 14                   | 18                    | 0.07               | 6           |
| Sandham JD            | Pulmonary artery catheter                   | 1,996       | 22                   | 28                    | 0.01               | 6           |
| van Nieuwenhoven CA   | Semirecumbent position                      | 221         | 4                    | 5                     | 0.01               | 6           |
| Van den Bergh G       | Intensive insulin therapy                   | 1,200       | 17                   | 25                    | 0.01               | 6           |
| Arabi YM              | Intensive insulin therapy                   | 523         | 8                    | 10                    | 0.01               | 6           |
| Brunkhorst FM         | Insulin therapy and pentastarch resuscitation | 537         | 15                   | 20                    | 0.02               | 4           |
| De La Rosa Gdel C     | Strict glycemic control                     | 504         | 11                   | 16                    | 0.02               | 6           |
| Kalfon P              | Intensive insulin therapy                   | 2,666       | 25                   | 35                    | 0.01               | 6           |
| Preiser JC            | Intensive insulin therapy                   | 1,101       | 15                   | 19                    | 0.01               | 6           |
| Augustine JJ          | Continuous vs intermittent dialysis         | 80          | 11                   | 16                    | 0.13               | 5           |
| Mehta RL              | CRRT vs IHD                                 | 164         | 13                   | 15                    | 0.07               | 6           |
| Uehlinger DE          | CRRT vs IHD                                 | 125         | 10                   | 15                    | 0.08               | 6           |
| Vinsoneau C           | CRRT vs IHD                                 | 359         | 16                   | 22                    | 0.05               | 6           |
| Bellomo R             | Intensity of CRRT                           | 1,464       | 39                   | 44                    | 0.02               | 5           |
| Palevsky PM           | Intensity of CRRT                           | 1,124       | 22                   | 30                    | 0.02               | 6           |
| Gaudry S              | Timing of RRT                               | 619         | 21                   | 26                    | 0.04               | 6           |
| Zarbock A             | Timing of RRT                               | 231         | 5                    | 9                     | 0.02               | 6           |
| Cook D                | Dalteparin vs unfractionated heparin        | 3,746       | 15                   | 21                    | 0.004              | 6           |
| Harvey SE             | Enteral vs parenteral nutrition             | 2,388       | 31                   | 40                    | 0.01               | 6           |
| Doig GS               | Early parenteral nutrition                  | 1,372       | 22                   | 27                    | 0.01               | 7.5         |
| Arabi YM              | Permissive underfeeding                     | 894         | 20                   | 25                    | 0.02               | 6           |
| Singh G               | Postoperative enteral feeding               | 43          | 7                    | 8                     | 0.16               | 4           |
| Petros S              | Hypo vs normocaloric                        | 100         | 1                    | 2                     | 0.02               | 6           |
| Reignier J            | Not monitoring gastric residual volume      | 449         | 13                   | 16                    | 0.02               | 6           |
| Valenta J             | High-dose selenium                          | 150         | 7                    | 9                     | 0.04               | 4           |
| Caparrós T            | High-protein diet enriched with arginine, fiber, antioxidant | 220         | 4                    | 7                     | 0.03               | 7.5         |
| Kieft H               | Immunonutrition                             | 597         | 17                   | 26                    | 0.03               | 8           |
| Grau T                | Immunonutrition                             | 127         | 8                    | 10                    | 0.07               | 8           |
| Galbán C              | Immune-enhancing diet                       | 176         | 1                    | 6                     | 0.03               | 6           |
| Puskarich MA          | LCarnitine                                  | 31          | 5                    | 6                     | 0.19               | 8           |
| Young P               | Buffered crystalloid vs saline              | 2,092       | 21                   | 28                    | 0.01               | 8           |
| Finfer S              | Albumin vs saline                            | 6,997       | 65                   | 80                    | 0.09               | 8           |

EGDT, early goal-directed therapy; HFOV, high-frequency oscillating ventilation; NIV, noninvasive ventilation; PEEP, positive end-expiratory pressure; PAC, pulmonary artery catheter; CRRT, continuous renal replacement therapy; IHD, intermittent hemodialysis

**DISCUSSION**

This retrospective analysis of evidences that formulated SSG found that the guidelines are based on highly robust RCTs with statistically insignificant results and on some moderately robust RCTs with statistically significant results. The median sample size was larger in RCTs having nonsignificant statistical results. Fl has been evaluated on studies of anticancer medicines, heart failure, anesthesiology, and several other areas of biomedical field in order to assess the robustness of findings amid concern over the reproducibility of research. A retrospective analysis calculated a median Fl of 56 RCTs in critical care medicine reporting mortality. The median Fl was 2 with an interquartile range (IQR) of 1 to 35. Similar to our study, several clinical guidelines were subjected to Fl analysis. An analysis of 32 RCTs included in the American College of Gastroenterology Guidelines of Crohn’s disease reported a median Fl of 3. An analysis of 21 RCTs that were used to support treatment recommendations in the 2016 Chest
Robustness of Surviving Sepsis Guidelines

Table 3: Subgroup analysis of RCTs according to domains they dealt with

| Subject                          | Studies with significant results | Studies with non-significant results | FI   | FQ   | RFI  | RFQ  |
|----------------------------------|----------------------------------|--------------------------------------|------|------|------|------|
| EGDT/GDT                         | 1                                | 7                                    | 4    | 0.01 | 6    | 0.02 |
| Vasopressors/inotropes           | 1                                | 5                                    | 5    | 0.15 | 10   | 0.02 |
| Infection                        | 2                                | 6                                    | 5    | 0.0045 | 12 | 0.03 |
| Ventilation, ECMO, and others related to oxygenation | 7                                | 8                                    | 4    | 0.01 | 12   | 0.03 |
| Nutrition                        | 1                                | 12                                   | 3    | 0.09 | 7.5  | 0.03 |
| Steroids                         | 1                                | 5                                    | 7    | 0.17 | 11   | 0.05 |
| Adjunct therapy                  | 1                                | 6                                    | 15   | 0.008 | 17.5 | 0.025 |
| Insulin therapy                  | 2                                | 6                                    | 8    | 0.002 | 15   | 0.01 |
| Transfusion                      | —                                | 3                                    | —    | —    | 6    | 0.02 |
| Anticoagulant/DVT prophylaxis     | —                                | 1                                    | —    | —    | 15   | 0.004 |
| Renal replacement therapy        | —                                | 8                                    | —    | —    | 16   | 0.03 |
| Patient position                 | 1                                | 1                                    | 3    | 0.02 | 4    | 0.03 |
| Pulmonary artery catheter        | —                                | 5                                    | —    | —    | 21   | 0.03 |
| Intravenous fluids               | —                                | 3                                    | —    | —    | 40   | 0.03 |
| End-of-life care                 | 1                                | 0                                    | 6    | 0.01 | —    | —    |
| Physical therapy                 | 1                                | 0                                    | 3    | 0.02 | —    | —    |
| Others                           | 1                                | 1                                    | 30   | 0.2  | 8    | 0.02 |

Guideline and Expert Panel Report on Antithrombotic Therapy for VTE Disease found a median FI score of 5 (1–9).26 Another study of 35 RCTs in the 2017 diabetes treatment guidelines reported that the median FI score was 16 (4–29).27 Analysis of 25 RCTs in heart failure reported a median FI score of 26 (0–118).16 Compared to these guidelines, RCTs of SSG had moderate robustness having a median FI of 5.5. Although there is no established cutoff value for FI or RFI as being robust or fragile, it is reasonable to postulate that the higher the value, the more “confidence” is on the possibility of the observed result to be robust. Studies that evaluated RCTs of various specialties reported median FI in the range of 2 to 26.13–15,17,24 A study calculated FI of 399 RCTs published in NEJM, JAMA, The Lancet, BMJ, and Annals of Internal Medicine. Median FI was 8 with an IQR of 0 to 10.11 The concept of RFI is relatively new. A recent study that analyzed 167 RCTs with statistically insignificant results that were published in NEJM, The Lancet, and JAMA reported a median RFI of 8 (5–13) at a p value of 0.05, which was lower than the median RFI of survival sepsis guidelines 2016.8

The FI and RFI are powerful and intuitive statistical concepts. They provide a useful additional tool for clinicians to use in assessing the treatment effect on patient outcomes. FI or RFI can help researchers to identify trials that are at risk of being overturned by future studies and avoiding overestimation of the significance of RCT results. However, looking at FI or RFI, it has been kept in consideration that many factors may influence them; of which, sample size, event rates, significant level, and statistical methods of association are important.28

The initial SSC guidelines were first published in 2004.29 Since then, it has changed clinical behavior, improved quality of care, and decreased mortality in patients with severe sepsis and septic shock. The studies demonstrated that increased compliance was associated with a 25% relative risk reduction in mortality rate.30 To our knowledge, analysis of FI and RFI of RCTs of these landmark guidelines was not done before. The present study may be first of its kind to assess the robustness of evidences that have shaped the guidelines. Previous studies appraising various clinical guidelines focused only on RCTs with significant results. Our study for the first time analyzed guidelines in regard to its RCTs with statistically insignificant results and also demonstrated that in these guidelines, RCTs with insignificant results are more robust than RCTs with statistically significant results.

Like any other statistical parameters, FI and RFI have also their own limitations. It can be used only to RCTs with dichotomous outcomes and 1:1 parallel study. RCTs with continuous outcomes cannot be evaluated. They do not account for the time at which events occurred which is a very important consideration, especially in oncological research.31 FI alone does not convey a measure of precision so it has to be read in conjunction with the p value, sample size, CI, and number lost to follow-up. Because of these limitations, the present study could not analyze less than half of the RCTs included in SSG.

This is to be noted that clinical decision about the effectiveness of harm of an intervention should not be merely based on the statistical significance or lack of it.32 Rather, it should be based on the magnitude of the treatment effect.32 The statistical significance merely tries to quantify the probability of observing the reported effect size. FI and RFI do not quantify the treatment effect; rather, they can be used to understand the fragility of the probability of the treatment effect reported.

This analysis of 100 RCTs that contributed to SSG found a median FI of 5.5 and a median RFI of 13. Most RCTs had statistically nonsignificant results, and they are more robust than statistically significant studies.

Contribution of Authors
Study design: NSC, SKD, PS, SD and SR; data analysis, acquisition, and interpretation: NSC, SKD, SD and PS; quality assessment: PS; drafting of manuscript: NSC, SKD, PS, and SR.
REFERENCES

1. Dahiru T. P-value, a true test of statistical significance? A cautionary note. Ann Ib Postgrad Med 2008;6(1):21–26. DOI: 10.4314/ajpm.v6i1.64038.

2. Nuzzo R. Scientific method: statistical errors. Nature 2014;506(7487):150–152. DOI: 10.1038/506150a.

3. Bertolaccini L, Viti A, Terzi A. Are the fallacies of the P value ref e r e n c e s fragility of statistically significant findings from randomized trials in the medical literature. Br J Anaesth 2018;120(5):935–941. DOI: 10.1093/bja/aey521.

4. Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The fragility of statistically significant results in otolaryngology randomized trials. Am J Otolaryngol 2019;40(1):61–66. DOI: 10.1016/j.aearlgy.2018.10.011.

5. Wayant C, Scott J, Vassar M. Evaluation of lowering the P value threshold for statistical significance from .05 to .005 in previously published randomized clinical trials in major medical journals. JAMA 2018;320(17):1813–1815. DOI: 10.1001/jama.2018.12288.

6. Condon TM, Sexton RW, Wells AJ, To MS. The weakness of fragility of statistically significant findings in randomized trials that support diabetes treatment guidelines for VTE? Chest. 2018;154(3):512–520. DOI: 10.1016/j.chest.2018.01.031.

7. Majeed M, Agrawal R, Attar BM, Kamal S, Patel P, Omar YA, et al. The fragility of statistically significant findings in otolaryngology randomized trials. Am J Otolaryngol 2019;40(1):61–66. DOI: 10.1016/j.aearlgy.2018.10.011.

8. Ayant C, Scott J, Vassar M. Evaluation of lowering the P value threshold for statistical significance from .05 to .005 in previously published randomized clinical trials in major medical journals. JAMA 2018;320(17):1813–1815. DOI: 10.1001/jama.2018.12288.

9. Halsey LG. The reign of the p-value is over: what alternative threshold for statistical significance from .05 to .005 in previously published randomized clinical trials in major medical journals. JAMA 2018;320(17):1813–1815. DOI: 10.1001/jama.2018.12288.

10. Oremus M, Wolfson C, Perrault A, Demers L, Momoli F, Moride Y. Interater reliability of the modified Jadad quality scale for systematic reviews of Alzheimer’s disease drug trials. Dement Geriatr Cogn Disord 2001;12:232–236. DOI: 10.1159/000051263.

11. Lin L. Factors that impact fragility index and their visualizations. Ophthalmology 2018;125(5):642–648. DOI: 10.1016/j.ophtha.2017.11.015.

12. Ayanti C, Scott J, Vassar M. Evaluation of lowering the P value threshold for statistical significance from .05 to .005 in previously published randomized clinical trials in major medical journals. JAMA 2018;320(17):1813–1815. DOI: 10.1001/jama.2018.12288.

13. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med 2019;45(9):1604–1664. DOI: 10.1007/s00134-019-05311-w.

14. 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.jcvi.2017.11.015.

15. Sumit Ray, Saurabh K Das, Nang S Choupoo, Priyam Saikia, Samarjit Dey, Indian Journal of Critical Care Medicine, Volume 25 Issue 7 (July 2021). 779

Orcid
Nang S Choupoo https://orcid.org/0000-0001-6270-3981
Saurabh K Das https://orcid.org/0000-0001-7798-4528
Priyam Saikia https://orcid.org/0000-0001-6608-484X
Samarjit Dey https://orcid.org/0000-0001-8211-253X
Sumit Ray https://orcid.org/0000-0001-5192-4711