Anaesthesia 2019, 74, 1057–1060 doi:10.1111/anae.14656

Review Article

The Fragility Index in peri-operative randomised trials that reported significant mortality effects in adults

L. Bertaggia,[1] **M. Baiardo Redaelli,**[2] **R. Lembo,**[3] **C. Sartini,**[2] **R. Cuffaro,**[2] **F. Corrao,**[2] **A. Zangrillo,**[2,4] **G. Landoni**[2,4] and **R. Bellomo**[5,6,7]

1 Medical Student, 2 MD, 3 Biostatistician, Department of Anaesthesia and Intensive Care, IRCCS San Raffaele Scientific Institute, Milan, Italy
4 Professor, Vita-Salute San Raffaele University, Milan, Italy
5 Professor, Faculty of Medicine, University of Melbourne, Melbourne, Vic., Australia
6 Professor, Australian and Intensive Care Research Centre, School of Public Health and Preventive Medicine, Monash University, Melbourne, Vic., Australia
7 Professor, Department of Intensive Care, Austin Hospital, Melbourne, Vic., Australia

Summary

More than one million peri-operative patients die each year. Thus, small improvements in peri-operative care may save thousands of lives. However, clinicians need confidence in the robustness of trial findings. The Fragility Index may complement frequentist analysis and provide quantitative assessment of robustness. We searched MEDLINE for peri-operative critical care randomised controlled trials that reported a statistically significant difference in mortality. We identified 46 trials with 37,347 participants. The median (IQR [range]) Fragility Index was 2 (1–3 [0–49]). Eleven trials had a Fragility Index of zero (changing from the Chi-square test to Fisher's exact test removed significance) and seven trials had a Fragility Index of 1. Only 23/46 trials had a Fragility Index greater than the number of patients lost to follow-up. There was a strong positive correlation between the Fragility Index and: the number of participants, $R^2 = 0.97$, $p < 0.0001$; the number of centres that recruited participants, $R^2 = 0.96$, $p < 0.0001$; the number of nations that recruited participants, $R^2 = 0.93$, $p < 0.0001$; and the number of deaths, $R^2 = 0.97$, $p < 0.0001$. As measured by the Fragility Index, the effect of peri-operative interventions on mortality in individual randomised controlled trials are not robust.

Correspondence to: G. Landoni
Email: landoni.giovanni@hsr.it
Accepted: 8 March 2019
Keywords: anaesthesia; fragility Index; mortality; peri-operative; randomised controlled trial
Twitter: @giovannilandoni

Introduction

Worldwide, approximately 230 million patients have major operations each year, more than one million of whom die within 30 days [1]. Small reductions in peri-operative mortality would, therefore, save thousands of lives. Despite its importance, there is little evidence to support peri-operative medical practice, possibly due to the complexity of the clinical setting [2, 3]. The use of p values, generated by frequentist analyses, to assess evidence has recently been questioned [4–8]: statistically significant mortality effects are often based on a small number of events, with p values sensitive to single events [9]. The Fragility Index is the number of participants without events (in the group with fewest events) who would have to experience an event to increase p from < 0.05 to ≥ 0.05 [10–13].

We aimed to calculate the Fragility Index of all peri-operative randomised controlled trials that reported a significant effect of an intervention on mortality and to assess its association with trial characteristics.

Methods

We did not seek ethical approval for this study. We searched PubMed and MEDLINE on 25 November 2018 for randomised controlled trials with adult participants (≥ 18 years old) that tested a non-surgical intervention within 24 h of surgery in at least half of participants and that reported an effect on mortality ($p < 0.05$) at one or more time-points (see also Supporting Information Table S1). The Fragility Index was originally only calculated for trials that: had two parallel arms or a two-by-two factorial design; allocated participants in a 1:1 ratio to treatment and control; and that had a dichotomous outcome, or a time-to-event outcome, reported as significant in the abstract [12]. We extended these criteria to trials that: had three parallel arms, for which we combined two groups against the group with the least events (see also Supporting Information Table S2); allocated participants in a ratio other than 1:1; and reported significant effects on survival anywhere in the manuscript. We performed a secondary analysis excluding trials that did not fulfil the original criteria.

We excluded trials that: did not randomly allocate participants; included children; and reported significant mortality effects for sub-groups only or as part of a composite outcome. Two investigators independently assessed trial eligibility; a third investigator resolved disagreements. The Fragility Index is the number of events required to reach a p value of 0.05 or more, calculated with Fisher's exact test, when added to the group with the fewest events. A Fragility Index of zero is possible if Fisher's exact text calculated $p > 0.05$ with the same data from which the Chi-square calculated $p < 0.05$ (if that is what the authors used to claim statistical significance). We used Pearson's correlation test, with Bonferroni correction for continuous variables and median test for dichotomous data, to assess the relationship between the Fragility Index and several trial characteristics: the number of allocated participants; the number of participating centres; the reported p value; blinding; the number needed to treat or to harm; the year of publication; whether the intervention increased or decreased mortality; the number of deaths; whether mortality was the primary outcome; and whether the intervention was in the operating room, or in intensive care. We used a logistic regression model with stepwise selection to identify independent associations with the Fragility Index. We entered variables into the model if univariate $p < 0.05$. Collinearity and overfitting were assessed using a stepwise regression model and Pearson's correlation test. We analysed mortality the last time it was reported significant. When the number of patients lost to follow-up at the time-point of interest was not clearly stated we considered the patients not reported either as dead or alive equal to the number of patients lost to follow-up.

Results

We analysed 46 trials with 37,347 participants, recruited by 430 centres (Fig. 1 and also see Supporting Information Table S3). Seven trials did not meet the original Fragility Index inclusion criteria (see also Supporting Information Tables S3 and S4 and Supporting Information References [2, 3, 29, 32, 54, 55, 62]). Participants were recruited by one centre in 28 trials and by multiple centres in 18 trials, four of which were multinational. The median (IQR [range]) number of participants was 171 (99–409 [37–20,211]). The intervention was intra-operative for 24,389 participants (65%) and before or after surgery for the rest.

The median (IQR [range]) Fragility Index was 2 (1–3 [0–49]) and 2 (1 – 3 [0–49]) without the seven trials that did not meet the original inclusion criteria. The Fragility Index was zero for 11 trials and it was one for seven trials. The number of participants lost to follow-up exceeded the Fragility Index in 23/46 trials: the median (IQR [range]) Fragility Index of these trials was significantly greater than the other trials, 2 (1–3 [1–6]) vs. 1 (0–2 [0–49]), $p = 0.03$. The Fragility Index correlated with: the number of trial participants, $R^2 = 0.97$, $p < 0.0001$; the number of recruiting centres, $R^2 = 0.96$, $p < 0.0001$; the number of nations, $R^2 = 0.93$, $p < 0.0001$; and the number of deaths, $R^2 = 0.97$, $p < 0.0001$. There was no correlation with: the year of publication, $R^2 = 0.02$, $p = 1$; the reported p value, $R^2 = -0.45$, $p = 0.12$; and the number needed to treat or harm, $R^2 = 0.43$, $p = 0.08$. There was no association of the median (IQR [range]) Fragility Index with whether the intervention was unblinded (27 trials) or blinded (19 trials), 2 (0–2 [0–6]) vs. 2 (1–3 [0–49]), respectively, $p = 0.051$. There was no association of the median (IQR [range]) Fragility Index with whether the primary outcome was mortality (29 trials) or not (17 trials), 2 (0–2 [0–8]) vs. 2 (1–3 [0–49]), respectively, $p = 0.08$. There was no association of the median (IQR [range]) Fragility Index with whether the intervention increased mortality (7 trials) or decreased mortality (39 trials), 2 (1–4 [0–5]) vs. 2 (0–3 [0–49]), respectively, $p = 0.4$. There was no association of the median (IQR [range]) Fragility Index with whether the population was all surgical (17 trials) or not (29 trials), 1 (1–2 [0–49]) vs. 2 (0–3 [0–8]), respectively, $p = 0.3$.

The Fragility Index was independently associated with the number of allocated participants and the number of reported deaths on multivariate analyses (see also

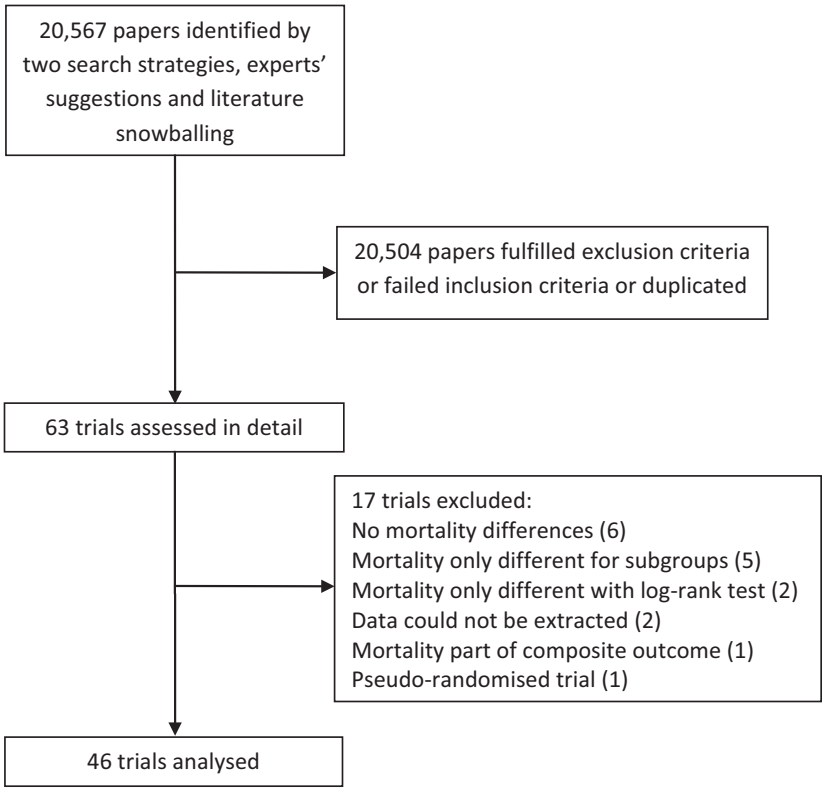

Figure 1 Flow chart of the study selection.

Supporting Information Table S5a). Univariate analyses without seven trials that did not meet the original Fragility Index criteria gave similar results, although the number needed to treat or harm in a trial was also associated with the Fragility Index, $R^2 = 0.50$, $p = 0.039$. There was also an association of median (IQR [range]) Fragility Index with whether mortality was the primary outcome (16 RCTs), 3 (2–4 [0–49]) vs. 2 (0–2 [0–5]), $p = 0.017$. Without these seven trials the Fragility Index was independently associated with the number of reported deaths on multivariate analyses (see also Supporting Information Table S5b).

Discussion

The median Fragility Index of interventions that affected peri-operative mortality was low, with a Fragility Index of zero in a quarter of trials, whereas in half of the trials the number of participants lost to follow-up was equal to, or exceeded, the trial's Fragility Index. There were strong positive correlations between the Fragility Index and some trial characteristics. Our findings were similar when we excluded trials that did not meet the original Fragility Index inclusion criteria.

A recent study reported a median Fragility Index of 2 in critical care trials with the number of participants lost to follow-up exceeding the Fragility Index in 7/56 trials, similar to our findings [9]. Trials published in five journals with a high impact factor had a higher median Fragility Index of 8, with only one quarter of the trials having a Fragility Index less than 4 [12].

Our systematic literature search without date limitation was a strength of our study, although it was limited to one database. We limited trials to those that reported a significant effect on mortality, which is not the only important patient-centred outcome. We focused on a specific group of patients, which decreases heterogeneity and confounding. There were other potential risks of bias that may have influenced the results. Some authors may argue that the Fragility Index carries similar information to the p value, but we think that the Fragility Index conveys the uncertainty of a statistically significant result in a way that is more accessible than the combination of p value and 95%CI.

In summary, the statistical significance of interventions that changed peri-operative mortality were often sensitive to only a few more participants dying in one of the groups: in almost a quarter of trials significance was lost by changing the statistical test. In the majority of trials the number of patients lost to follow-up was greater than the number of deaths that would have negated the statistically significant

result. The addition of the Fragility Index to p values and 95%CI might improve the understanding that categorical outcomes are often uncertain even when statistically significant.

Acknowledgements

We thank Maieutics Foundation for their support with data analysis and G. Giardina and G. Dalessandro for their help with data entry. No competing interests declared.

References

1. Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008; **372**: 139–44.
2. Landoni G, Comis M, Conte M, et al. Mortality in multi center critical care trials: an analysis of interventions with a significant effect. *Critical Care Medicine* 2015; **43**: 1559–68.
3. Landoni G, Pisano A, Lomivorotov V, et al. Randomized evidence for reduction of perioperative mortality: an updated consensus process. *Journal of Cardiothoracic and Vascular Anesthesia* 2017; **31**: 719–30.
4. Sterne JA, Davey SG. Sifting the evidence-what's wrong with significance tests? *British Medical Journal* 2001; **322**: 226–31.
5. Rozeboom WW. The fallacy of the null-hypothesis significance test. *Psychological Bulletin* 1960; **57**: 416–28.
6. Rothman KJ. Significance questing. *Annals of Internal Medicine* 1986; **105**: 445–7.
7. Goodman SN. Toward evidence-based medical statistics. 1: the P value fallacy. *Annals of Internal Medicine* 1999; **130**: 995–1004.
8. McIlroy D. Seduced by a P-value. *Anaesthesia and Intensive Care* 2014; **42**: 551–4.
9. Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The Fragility Index in multi center randomized controlled critical care trials. *Critical Care Medicine* 2016; **44**: 1278–84.
10. Feinstein AR. The Unit Fragility Index: an additional appraisal of "statistical significance" for a contrast of two proportions. *Journal of Clinical Epidemiology* 1990; **43**: 201–9.
11. Walter SD. Statistical significance and fragility criteria for assessing a difference of two proportions. *Journal of Clinical Epidemiology* 1991; **44**: 1373–8.
12. 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. *Journal of Clinical Epidemiology* 2014; **67**: 622–8.
13. Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine Journal* 2015; **15**: 2188–97.

Supporting Information

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

Table S1. Search strategy.

Table S2. Management of trials with three groups.

Table S3. Trial characteristics.

Table S4. Exclusions.

Table S5. Multivariate analyses.

Data S1. References of analysed trials.