BJA

British Journal of Anaesthesia, ■ (■): 1–7 (2018)

doi: 10.1016/j.bja.2018.01.012
Advance Access Publication Date: xxx
Review Article

REVIEW ARTICLE

The fragility of statistically significant findings in randomised controlled anaesthesiology trials: systematic search of the medical literature

G. Mazzinari[1,*], L. Ball[2], A. Serpa Neto[3,4], C. L. Errando[5], A. M. Dondorp[6], L. D. Bos[4], M. G. de Abreu[7], P. Pelosi[2] and M. J. Schultz[4,6]

[1]Research Group in Perioperative Medicine, Hospital Universitari i Politècnic La Fe, Valencia, Spain, [2]Department of Surgical Sciences and Integrated Diagnostics, Policlinico San Martino Hospital – IRCCS for Oncology, Genoa, Italy, [3]Department of Critical Care Medicine, Hospital Israelita Albert Einstein, São Paulo, Brazil, [4]Department of Intensive Care and Laboratory of Experimental Intensive Care and Anesthesiology (L·E·I·C·A), Academic Medical Center, Amsterdam, The Netherlands, [5]Department of Anesthesiology, Resuscitation and Pain Treatment, Consorcio Hospital General Universitario de Valencia, Valencia, Spain, [6]Mahidol Oxford Research Unit (MORU), Mahidol University, Bangkok, Thailand and [7]Department of Anesthesiology and Intensive Care Therapy, Pulmonary Engineering Group, Technische Universität Dresden, Dresden, Germany

*Corresponding author. E-mail: gmazzinari@gmail.com.

Abstract

The fragility index (FI), the number of events the statistical significance a result depends on, and the number of patients lost to follow-up are important parameters for interpreting randomised clinical trial results. We evaluated these two parameters in randomised controlled trials in anaesthesiology. For this, we performed a systematic search of the medical literature, seeking articles reporting on anaesthesiology trials with a statistically significant difference in the primary outcome and published in the top five general medicine journals, or the top 15 anaesthesiology journals. We restricted the analysis to trials reporting clinically important primary outcome measures. The search identified 139 articles, 35 published in general medicine journals and 104 in an anaesthesiology journals. The median (inter-quartile range) sample size was 150 (70–300) patients. The FI was 4 (2–17) and 3 (2–7), and the number of patients lost to follow-up was 0 (0–18) and 0 (0–6) patients in trials published in general medicine and anaesthesiology journals, respectively. The number of patients lost to follow-up exceeded the FI in 41 and 27% in trials in general medicine journals and anaesthesiology journals, respectively. The FI positively correlated with sample size and number of primary outcome events, and negatively correlated with the reported *P*-values. The results of this systematic review suggest that statistically significant differences in randomised controlled anaesthesiology trials are regularly fragile, implying that the primary outcome status of patients lost to follow-up could possibly have changed the reported effect.

Keywords: anaesthesiology; lost to follow-up; randomised controlled trials; research methodology; statistical significance

Editorial decision: January 14, 2018; **Accepted:** January 14, 2018

Editor's key points

- The authors examined the fragility index (FI) and the number of patients lost to follow-up in randomised controlled trials with major clinical endpoints as the primary outcome in anaesthesiology.
- The FI was correlated positively with sample size and number of primary outcome events, and negatively with reported *P*-values. The number of patients lost to follow-up exceeded FI in a third of the trials.
- The authors concluded that statistically significant differences in randomised controlled anaesthesiology trials are regularly fragile, suggesting that primary outcomes in patients lost to follow-up could have influenced findings.

The fragility index (FI) of a randomised controlled trial is the number of patients in the randomisation group with the fewest primary outcome events, whose status would have to change from 'non-event' to 'event' to change a statistically significant difference between treatment arms to a non-significant difference.[1] For example, in a trial with an FI of 20, as many as 20 additional patients with an event would be needed to render a statistically significant difference non-significant. In contrast, an FI of 1 implies that only one patient changed to the alternative outcome status would change the overall result to a non-significant difference between treatment arms. Thus, the lower the FI the more 'fragile' is the statistical significance finding of a trial.

The number of patients lost to follow-up is defined as the number of patients in whom the status of the primary outcome remains unreported no matter the cause. Loss to follow-up could thus cause bias, in particular when the reasons for loss to follow-up are associated with the likelihood of occurrence of the primary outcome.[2] The number of patients lost to follow-up adds to the concept of FI. For instance, in a trial with an FI of 5 in which 10 patients are lost to follow-up, the difference between treatment arms would change to non-significant if five or more of these patients would have experienced the primary outcome event. In other words, if the FI is lower than the number of patients lost to follow-up, the statistical significance finding of a trial could be even more 'fragile'.

Previous investigations reporting on FI and number of patients lost to follow-up in randomised controlled trials in the domains of general medicine,[1] spinal surgery,[3] intensive care medicine,[4] sports medicine,[5] and recently also cardiology[6] not only show that the FI is frequently low, but also that it is common that the number of patients lost to follow-up exceed the FI. These investigations also show that statistically significant findings are more fragile when trials are smaller in size or have a lower number of primary outcome events. As randomised controlled trials in the field of anaesthesiology are frequently small, and often have small numbers of primary outcome events, we hypothesised that statistically significant findings in these trials suffer from a comparable 'fragility' to those in trials in other domains of medicine. To test this hypothesis, we performed a systematic search of the medical literature, seeking for published randomised controlled anaesthesiology trials to calculate the FI. We compared the FI with the number of patients lost to follow-up, and identified which factors were associated with the FI.

Methods

Search strategy

We performed a systematic search in PubMed/MEDLINE for randomised controlled anaesthesiology trials reporting a statistically significant difference between study arms with regard to the primary outcome. While we did not pre-publish the study protocol, we followed a strict pre-defined plan for both the search and the analysis.

We did not perform a formal power calculation. We set the boundaries of the systematic search rather pragmatically. First, we restricted the search to trials reported in the top five general medicine journals or the top 15 anaesthesiology journals, based on their impact factors at the moment we performed the actual search. We assumed that trials reported in these journals would be of high quality, and would contain sufficient information to calculate for example the number of patients lost to follow up. We also restricted the time window to the past 25 yrs (i.e. published since 1991), for the very same reasons. As we updated the search shortly before finalising this paper, we ended up with a time frame of 26 yrs. A complete list of search criteria and journals is presented in the Supplementary material (Table S1 and PubMed full search criteria).

Two reviewers (G.M. and L.B.) independently scanned all articles identified by the search for relevance by reading the title and abstract. For potentially relevant articles, the full text was obtained. In case of disagreement, consensus between the two reviewers was sought. Reference lists of initially selected articles, as well as related reviews and meta-analyses, were searched for additional potentially relevant articles.

Selection of studies

Articles were selected when the following criteria were met: 1) randomised controlled trial in humans, 2) study in the field of perioperative anaesthesiology, performed in operation theatres, with outcomes directly related to perioperative management, 3) reporting a statistically significant difference with regard to the primary outcome, and 4) primary outcome of the study was a 'major clinical endpoint' (see below for definitions). Trials were excluded if: 1) trial design was not a two parallel-arm or a two by two factorial randomised controlled trial (factorial meaning that the effects of two independent interventions were assessed in one single trial); 2) not using 1:1 randomised allocation; and 3) if not reporting on a dichotomic primary outcome, or when it was not possible to dichotomise time-to-event outcomes. In addition, trials using a quasi-or a non-randomised methodology were excluded, as were trials performed outside the operation room. In addition, trials with inaccuracies in the reported number of recruited patients were excluded, which was revealed when the number of patients with a primary outcome event *plus* the number of patients without a primary outcome event *plus* the number of patients lost to follow-up did not match the reported number of recruited patients, as well as trials stopped before reaching full recruitment because of safety concerns about the intervention tested.

As mentioned above, we restricted the analysis to trials that used a primary outcome measure that occurred in the intra- or postoperative phase and could be considered to be clinically important and relevant (i.e. a 'major clinical outcome')[7,8] (see Supplementary Table S2). The following outcome measures were included: mortality; delirium or