Fragility index and fragility quotient in randomized clinical trials

Marcos Vinicius Fernandes Garcia[1], Juliana Carvalho Ferreira[2,3,4], Pedro Caruso[2,3]

CLINICAL SCENARIO

In a randomized clinical trial (RCT) by Meyer et al.,[1] tenecteplase plus heparin was compared with placebo plus heparin in patients with pulmonary embolism. The primary outcome (death or hemodynamic decompensation) occurred in 13 of 506 patients (2.6%) in the intervention group as compared to 28 of 499 patients (5.6%) in the control group (OR = 0.44; 95% CI: 0.23-0.87; p = 0.02).

RCT ROBUSTNESS

RCTs are expensive and time consuming, and they generally have limited sample sizes; therefore, results can be dependent on few events. In the abovementioned RCT,[1] despite the large sample size, if only 3 more patients in the intervention group had experienced the outcome, the p-value would be greater than 0.05, which means that if 16 patients, rather than 13 patients, in the experimental group had experienced the primary outcome, the study would not be significant. This number indicating how many additional events in one of the groups would be required to turn a statistically significant trial into a statistically non-significant trial is called the fragility index (FI).

FI

The FI is calculated by changing the status of 1 patient in the group with the fewest number of events (control or experimental) from "non-event" (not experiencing the primary outcome) to "event" and then recalculating a two-sided Fisher's exact test until p becomes ≥ 0.05.[2] Table 1 illustrates the calculation of the FI for the abovementioned RCT.[1] Therefore, the FI is a measure of robustness of clinical trial results; the smaller the FI is, the less robust the trial is considered to be. Although the FI has no formal cutoff, it serves as an additional indicator of how easily the statistical significance of an RCT depends on a small number of events. Also, as a rule of thumb, if the number of patients lost to follow-up is greater than the FI, the trial should be considered less robust.

The fragility quotient (FQ) is the FI divided by the sample size, and a low FQ indicates a less robust trial. The FQ for the abovementioned RCT[1] would be 3/1,005 = 0.003, which is small and also indicates that the trial is not robust. FQ provides a way to assess the vulnerability of studies with regard to sample size, especially when sample sizes vary widely between studies addressing the same intervention.

FI USE AND LIMITATIONS

A large FI does not necessarily indicate a conclusive result, and a small FI does not indicate that the RCT results are trivial. There is no clear consensus on defining what a "fragile" study is, but FI and FQ can help clinicians make health decisions considering the fragility of RCT results.

Table 1. Example of calculation of the fragility index.[a]

Study sample (N = 1,005)	Death or hemodynamic decompensation	Neither death nor hemodynamic decompensation	p
Study outcome			0.02
Intervention group	13	493	
Control group	28	471	
First step of FI calculation			0.027*
Intervention group	14	492	
Control group	28	471	
Second step of FI calculation			0.043*
Intervention group	15	491	
Control group	28	471	
Third step of FI calculation			0.065*
Intervention group	16	490	
Control group	28	471	

FI: fragility index. [a]Adapted from Meyer et al.[1] *Fragility index steps and p-values were calculated using the "R package: fragility index."

1. Cleveland Clinic Foundation, Cleveland (OH) USA.
2. Divisão de Pneumologia, Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo (SP) Brasil.
3. Unidade de Terapia Intensiva, AC Camargo Cancer Center, São Paulo (SP) Brasil.
4. Methods in Epidemiologic, Clinical, and Operations Research-MECOR-program, American Thoracic Society/Asociación Latinoamericana del Tórax, Montevideo, Uruguay.

BY-NC © 2023 Sociedade Brasileira de Pneumologia e Tisiologia

The main limitation of the FI is that it applies only to RCTs with dichotomous outcomes. Another limitation is an FI equal to zero. While a trial uses chi-square analysis to calculate a p-value, FI is calculated using the Fisher's exact test, and therefore, an FI = 0 could occur in such trials when statistical significance was lost by simply changing the analysis from the chi-square test to the Fisher's exact test.

KEY MESSAGES

- The FI estimates the number of events needed to turn a statistically significant trial into non-significant. The smaller the FI is, the less robust the trial is.
- FI and FQ offer an alternative to the frequentist approach to RCT analysis and have been increasingly used in the critical appraisal of RCTs as an adjunctive tool for RCT interpretation.

REFERENCES

1. Meyer G, Vicaut E, Danays T, Agnelli G, Becattini C, Beyer-Westendorf J, et al. Fibrinolysis for patients with intermediate-risk pulmonary embolism. N Engl J Med. 2014;370(15):1402-1411. https://doi.org/10.1056/NEJMoa1302097
2. Tignanelli CJ, Napolitano LM. The Fragility Index in Randomized Clinical Trials as a Means of Optimizing Patient Care. JAMA Surg. 2019;154(1):74-79. https://doi.org/10.1001/jamasurg.2018.4318