CIRUGÍA ESPAÑOLA

www.elsevier.es/cirugia

Methodological letter

Fragility Index, Reverse Fragility Index, Fragility Quotients, and Susceptibility Index. Novel tools for assessing of randomized clinical trials☆

Indice de Fragilidad, Indice Inverso de Fragilidad, Cocientes de Fragilidad e Indice de susceptibilidad. Nuevas herramientas para la valoración de los ensayos clínicos aleatorizados

Carlos Placer-Galán, Elisabet Bollo Arocena*

Servicio de Cirugía General y Digestiva, Hospital Universitario Donostia, San Sebastián, Spain

Randomized clinical trials (RCTs) represent the paradigm of Evidence-Based Medicine (EBM) and are the primary methods used to evaluate the efficacy and safety of treatment studies. Additionally, they constitute the primary support for testing causality in interventions. Randomization and its concealment, both for researchers and trial participants, are the primary tools used to avoid biases. However, RCTs present some limitations, particularly those conducted during surgery.[1] One of the main problems in the correct interpretation of results is the loss of participants during the process. Different approaches such as protocol and intention-to-treat analyses have been proposed to overcome these difficulties. The external validity of these studies may be compromised by some factors, including professionals (varying degrees of competence in the aspect of study), participants (difficulty in selecting them), and intervention (most RCTs are conducted in third-level centers with experience that is difficult to extrapolate to other settings). In the field of surgery, RCTs present some more specific problems, including variations in surgical technique, learning curve, and comparison between study groups (patients with previous surgeries?) and difficulty in blinding the treatment performed (for both professionals and participants).[2]

Traditionally, *P*-values have been used to determine the statistical significance between two or more options in RCTs. This methodology has been criticized for not providing the strength of the associations and reducing their clinical importance. To overcome these deficiencies, most scientific journals require the incorporation of other statistics, such as relative risks, odds ratios, and confidence intervals, which help determine the true size of the effect achieved. For an average reader, these tests are not easy to interpret, and other metrics have been designed to facilitate the task. The number needed to treat (NNT) and number needed to harm (NNH) are two tools that provide a simpler and more practical response to interpreting the results of RCTs.

In recent years, new tools have appeared to evaluate RCTs more intuitively and make decisions about their fragility or robustness. These included the Fragility Index (FI), Reverse Fragility Index (RFI), Fragility Quotient, Reverse Fragility Quotient (FQ and RFQ), and Susceptibility Index (SI).

☆ Please cite this article as: Placer-Galán C, Arocena EB. Indice de Fragilidad, Indice Inverso de Fragilidad, Cocientes de Fragilidad e Indice de susceptibilidad. Nuevas herramientas para la valoración de los ensayos clínicos aleatorizados. Cir Esp. 2023.

* *Corresponding author.*
E-mail address: carlos943313762@gmail.com (C. Placer-Galán).

http://dx.doi.org/10.1016/j.cireng.2023.11.001

Fragility Index, Fragility Quotient and Susceptibility Index

FI is defined as the minimum number of outcomes of a trial that would need to be changed to reverse statistical significance.[3] This test can only be used in RCTs with a dichotomous outcome (death or no death, recurrence, or no recurrence) and a significant *P*-value ($P < .05$). In practice, FI is the number of "no events" in the experimental group with the smallest number of events that should be changed to "event" to change the *P*-value to >.05 without modifying the total number of participants (using Fisher's exact method).

Let us now examine an example. One trial compared two treatments: A (experimental group) and B (control group). The sample size of the study was 50 participants. In group A, there were 2 recurrences out of 25, and in group B, there were 9 recurrences out of 25 with a *P*-value of .037 (Fisher's exact test), which was statistically significant. If we increase the number of events in group A from two to three, the *P*-value would be .095, and therefore, significance would be lost. It can be concluded that the FI is 1. The smaller the FI value, the more fragile the study. A cutoff point defining the degree of fragility/robustness of a trial has not yet been defined, but studies with FI < 3 are usually considered fragile.

The sample size has a significant influence on the external validity of a randomized controlled trial, and the Fragility Quotient (FQ) has been defined by dividing the FI by the sample size.[4] In our example, it would be 1/50 (0.02). The interpretation is in terms of the number of patients per 100. In other words, in our example, a 2% change in the results of the study would rule out the superiority of Treatment A.

Given that FI is the minimum number of events that would change the significance, it is crucial to consider the value of lost patients during the study. Furthermore, if the number of patients lost is equal to or greater than the FI, the study can be considered fragile. This new concept has been incorporated into the Susceptibility Index.[5] Its calculation follows the following formula: SI = (number of lost patients–FI)/(number of lost patients). Higher values indicate that the results are highly susceptible to alteration by the number of patients lost during follow-up and facilitate comparison between studies.

Reverse fragility index and reverse fragility quotient

As already mentioned, FI can only be used in RCTs that have demonstrated the superiority of one treatment over another. However, what happens in trials in which statistical significance is not achieved ($P > .05$)? Recently, the RFI has been described, which is calculated in a similar way as the FI, subtracting events from the group with fewer events while adding "no events" to the same group, maintaining the total number of participants until the *P*-value was <.05.[6] Its interpretation is very similar to that of FI. A low RFI value indicated a weaker or more fragile trial. Its Rverse Fragility Quotient (RFQ) was also designed, which was calculated as the QF, that is, by dividing the RFI value by the sample size, and whose interpretation is the same as the fragility quotient.

An index of susceptibility has not yet been published or validated for studies with a *P*-value of >.05.

The development and validation of a susceptibility index for studies with a *P*-value >.05 has not yet been published.

Limitations and strengths of FI, RFI, FQ, RFQ

All indices discussed can only be used in studies with binary outcomes and a 1:1 design. Therefore, they cannot be used in trials that compare three or more treatments or continuous variables (e.g., pain scales). Additionally, they did not provide information on the clinical significance and importance of these differences.

However, these new tools are useful and intuitive measures that allow for the evaluation of RCTs and comparisons between them. Moreover, they can be applied to meta-analyses, where they can substantially change the conclusions.

How to the calculate simply

To perform these simple calculations, any spreadsheet (Excel, etc.) or statistical software (SPSS, STATA, SAS, R, etc.) can be used to construct a 2 × 2 contingency table (two study groups and two outcomes). The change in statistical significance was determined using Fisher's exact test. In cases where the authors of the study used $\chi 2$ and a *P*-value of <.05, sometimes the exact Fisher test results in a *P*-value of >.05. In these cases, the FI was 0.

For interested readers, there is a website that automatically calculates the FI (https://clincalc.com/Stats/FragilityIndex.aspx). Additionally, R software has a specific package for performing all the calculations described in this paper (R software version 4.2.2, R package Fragility Index).

Conflict of interest

The authors have no conflict of interest in this study.

REFERENCES

1. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ. 2002;324(7351):1448–51. http://dx.doi.org/10.1136/bmj.324.7351.1448.
2. Sancho JJ. Particularidades de los ensayos clínicos en el ámbito de la cirugía. Cir Esp. 2004;75:3–8. http://dx.doi.org/10.1016/S0009-739X(04)72264-2.
3. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, 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):622–8. http://dx.doi.org/10.1016/j.jclinepi.2013.10.019.
4. Forrester LA, McCormick KL, Bonsignore-Opp L, Tedesco LJ, Baranek ES, Jang ES, et al. Statistical fragility of surgical

clinical trials in orthopaedic trauma. J Am Acad Orthop Surg Glob Res Rev. 2021;5(11):e20.00197. http://dx.doi.org/10.5435/JAAOSGlobal-D-20-00197.
5. Majeed M, Agrawal R, Attar BM, Kamal S, Patel P, Omar YA, et al. Fragility index: how fragile is the data that support the American College of Gastroenterology guidelines for the management of Crohn's disease? Eur J Gastroenterol Hepatol. 2020;32(2):193–8. http://dx.doi.org/10.1097/MEG.0000000000001635.
6. Kahn MS, Fonarow GC, Friede T, Lateef N, Khan SU, Anker SD, et al. Application of the reverse fragility index to statistically nonsignificant randomized clinical trial results. JAMA Netw Open. 2020;3(8):e2012469. http://dx.doi.org/10.1001/jamanetworkopen.2020.12469.