A critique of the fragility index

In *The Lancet Oncology*, Joseph Del Paggio and Ian Tannock report on a retrospective analysis of the fragility indices of phase 3 trials used by the US Food and Drug Administration (FDA) to approve anticancer drugs.[1] The authors conclude that "many phase 3 randomised controlled trials supporting FDA-approved anticancer drugs have a low fragility index, challenging confidence for concluding their superiority over control treatments." Although this interpretation is interesting, we would like to draw attention to several points.

First, for many reasons, randomised controlled trials must be as efficient as possible, with the inclusion of only the necessary number of participants required to identify clinically meaningful differences in the trial's primary endpoint. Therefore, it is to be expected that the results of isolated randomised controlled trials will show fragility.

Second, it seems counterintuitive that Del Paggio and Tannock extracted and assessed secondary outcomes,[1] as trials are not adequately powered or designed to prove differences in secondary endpoints.[2]

Furthermore, a 2017 study of the fragility index found a correlation as high as 94% between p values and fragility indices;[3] thus, the idea that this will aid in the interpretation of trial data seems unlikely. Overall, when looking at the trials analysed, the results are easily interpretable simply by analysing the confidence intervals, as recommended by Cochrane.[4]

Finally, regulatory decisions require clinical knowledge and the analysis of more than one outcome per trial, a fact that is lost when the conclusions of these—mostly very positive—trials are put into question. Additionally, regulators acknowledge that the first available data will always be comparatively sparse, and only more trials, ideally pooled with meta-analyses, can provide a firm basis for the best clinical decisions. More is always better but, for regulators, fragile initial data are commonplace.

New metrics to analyse trial data periodically come into existence, although we can think of none that illuminate a previously unknown problem: most seem to be a slightly more convoluted way to reframe old problems, and almost none have staying power.

Metrics such as the fragility index, much like the p value, are easily confused as effect measures, raising the question of whether we really need a new metric to restate an old problem whose solution is only likely to be achievable through better clinical research education. We think not.

We declare no competing interests.

*Tiago Machado†, *Gonçalo S Duarte†, Nilza Gonçalves, Joaquim J Ferreira, João Costa*
gduarte@edu.ulisboa.pt

†Contributed equally.

Laboratory of Clinical Pharmacology and Therapeutics (TM, GSD, NG, JJF, JC) and Centre for Evidence-Based Medicine (JC), Faculty of Medicine, University of Lisbon, 1649-028 Lisbon, Portugal; Instituto de Medicina Molecular, Lisbon, Portugal (TM, GSD, NG, JJF, JC); and Campus Neurológico Sénior, Torres Vedras, Portugal (JJF)

1 Del Paggio JC, Tannock IF. The fragility of phase 3 trials supporting FDA-approved anticancer medicine: a retrospective analysis. *Lancet Oncol* 2019; **20:** 1065–69.
2 Friedman LM, Furberg C, DeMets DL. Fundamentals of clinical trials. New York: Springer, 2010.
3 Carter RE, McKie PM, Storlie CB. The fragility index: a p-value in sheep's clothing? *Eur Heart J* 2017; **38:** 346–48.
4 Schünemann HJ OA, Vist GE, Higgins JPT, et al. Interpreting results and drawing conclusions. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS, eds. Cochrane handbook for systematic reviews of interventions, version 5.2.0. Cochrane, 2017.