A critique of the fragility index

We commend Joseph Del Paggio and Ian Tannock on their study in *The Lancet Oncology*,[1] which aimed to assess the robustness of randomised clinical trials supporting the Food and Drug Administration approval of anticancer drugs. Such attempts to challenge the confidence in the superiority of anticancer drugs over control treatments are important, as significant differences in outcomes assessed using an arbitrary threshold (p<0·05) can be easily reversed with a change in designation of very few events, and might not reflect clinical benefit according to European Society for Medical Oncology or American Society of Clinical Oncology value scales.[2]

However, the study relies entirely on the fragility index—the minimal number of changes from non-events to events that would result in a loss of statistical significance—which does not account for the time at which events occurred. In contrast to studies in which the factor of time is less important (eg, success rates of a surgical procedure), time is often the most important factor in oncology trials. As an example, let us consider a well conducted trial testing a novel drug for metastatic pancreatic cancer which extends the median survival time by 5 years compared with the control, but with the same final proportion of events in both groups. Although this increased survival time would be considered a groundbreaking treatment for the disease, the resulting fragility index would be 0 (Fisher's exact test), deeming the trial to be fragile, with insufficient evidence for treatment superiority.

To provide a comprehensive solution, a possible alternative would be to extend the concept of the fragility index to survival analysis.[3] The survival fragility index is defined as the estimated number of individuals who had an event at the average exposure time of all individuals in the study, whose addition would result in a loss of statistical significance. Thus, the survival fragility index can provide a measure that both accounts for events over time and maintains all the advantages of the original fragility index.

We extracted time-to-event data from published Kaplan-Meier curves from the 17 trials analysed by Del Paggio and Tannock[1] using the DigitizeIt software and the method described by Wei and Royston.[4] This reverse-engineering strategy enables extrapolation of survival time data at the individual level, and succeeded in reproducing each dataset with minor differences between the published and inferred hazard ratios and events in individual groups.

Del Paggio and Tannock reported a median fragility index of 2 (IQR 0–27), with nine (53%) of 17 trials having an index of 2 or less, indicating poor robustness for most of the trials. However, we found the fragility index to correlate poorly with the survival fragility index (appendix). The median survival fragility index was 62 (IQR 21–71) and all trials had non-zero values. The disparity might suggest that use of the fragility index on time-to-event data is inappropriate in cases where the numbers of events in both groups are similar but the timing of events considerably differs.[5] Using this approach might lead to the false conclusion that such trials are fragile.

Challenging the confidence of clinical trial data is paramount to avoid approval of anticancer drugs with low clinical value and to verify the confidence in concluding superiority over control treatments. By incorporating time, the survival fragility index might provide a powerful measure to assess the strength of statistical conclusions.

We declare no competing interests.

*David Bomze, Tomer Meirson
davidbomze@mail.tau.ac.il

Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel (DB); and Drug Discovery Laboratory, The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel (TM)

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 Del Paggio JC, Sullivan R, Schrag D, et al. Delivery of meaningful cancer care: a retrospective cohort study assessing cost and benefit with the ASCO and ESMO frameworks. *Lancet Oncol* 2017; **18:** 887–94.

3 Johnson KW, Rappaport E, Khader S, Glicksberg BS, Dudley JT. fragilityindex: an R package for statistical fragility estimates in biomedicine. *bioRxiv* 2019; published online Feb 27. DOI:10.1101/562264 (preprint).

4 Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan-Meier curves. *Stata J* 2017; **17:** 786–802.

5 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. *J Clin Epidemiol* 2014; **67:** 622–28.

See **Online** for appendix