A critique of the fragility index

We read with interest the Article by Joseph Del Paggio and Ian Tannock[1] in *The Lancet Oncology* that explored the fragility of phase 3 trials supporting the approval of anticancer drugs. The authors should be congratulated for their attempts to quantify whether registration trials are robust.

Although the methods used by Del Paggio and Tannock mirror those used in the initial development of the fragility index,[2] these methods are difficult to apply to trials of anticancer medicines. The fragility index was designed to use dichotomous data, such as the presence or absence of an event of interest (eg, myocardial infarction). By contrast, anticancer drug trials use time-to-event data, in which the primary interest is the timing of events, rather than whether or not they have occurred. For example, disease progression occurs in almost all patients with metastatic solid tumours, but it is the duration of progression-free survival that is important. By using dichotomous data to calculate the fragility index, Del Paggio and Tannock have lost the ability to explore the timing of events, and their data reflect outcomes at a fixed timepoint.

Available data show that, compared to data derived using time-to-event analyses, calculating event probabilities at a fixed timepoint can result in both discordant magnitude of effect and loss of statistical power.[3] These observations are evident in the Article by Del Paggio and Tannock, in which six (35%) of 17 studies had a fragility index of 0, despite showing statistically significant effects based on time-to-event analyses of primary data. This finding questions the face validity of the reported results.

The Parmar Toolkit[4] can be used to estimate life tables from published studies, thereby allowing calculation of the fragility index using time-to-event data. By re-creating a time-to-event analysis, this method is analogous to the primary analyses of respective trials and can also address asymmetrical randomisation, a frequent design in cancer trials. Using the Parmar Toolkit, we calculated the fragility indices for the 17 studies reported by Del Paggio and Tannock (appendix). In general, we observed a higher fragility index using this method and did not observe any fragility index of 0. These results more closely reflect the primary analyses of the respective trials.

In conclusion, although Del Paggio and Tannock provide important data on the fragility of registration trials, readers should be aware of the limitations of the methods used and the uncertainty regarding the validity of their results.

*Alexandra Desnoyers,
Michelle B Nadler, Brooke E Wilson,
Eitan Amir
eitan.amir@uhn.ca

Division of Medical Oncology, Princess Margaret Cancer Centre and the University of Toronto, Toronto ON, M5G 2M9, Canada (AD, MBN, EA); University of New South Wales, Sydney, NSW, Australia (BEW); and Kinghorn Cancer Centre, Darlinghurst, NSW, Australia (BEW)

See **Online** for appendix

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. 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.

3. Michiels S, Piedbois P, Burdett S, et al. Meta-analysis when only the median survival times are known: a comparison with individual patient data results. *Int J Technol Assess Health Care* 2005; **21:** 119–25.

4. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998; **17:** 2815–34.