A critique of the fragility index

Authors' reply

We thank the readers for their constructive criticism of our analysis.[1] We were aware of the major limitation in using the fragility index for time-to-event data and discussed it in our Article. In the original report by Walsh and colleagues,[2] when the fragility index was applied to studies with time-to-event endpoints, linear regression led to "no material difference in the fragility index between time-to-event data and frequency data, which is consistent with the concept that most results are sensitive to the number of events in each group rather than the timing of the events".

We were not aware of previous applications of the fragility index to the field of oncology, nor alternative methods for fragility index analysis, and we followed the original methodology.[2] Given the reliance of oncology trials on time-to-event endpoints, the limitation of the Fisher's exact test-based fragility index becomes most apparent in the discordance between extreme fragility (ie, a fragility index of 0) and clinically meaningful differences within our dataset. We recognise that the fragility index will depend on the time at which a trial is analysed and reported.

To address the concern raised by Machado and colleagues with respect to the use of secondary endpoints, in our fragility index analysis, only one trial analysed in our dataset involved a secondary endpoint.[3] We agree that emphasis should not be placed on secondary endpoints when interpreting the results of clinical trials.[4]

Debate about the utility of the fragility index will continue, and we caution once again about its use in isolation when evaluating clinical trial data; nevertheless, low fragility index values might help to raise concerns about trial reliability, which must then be evaluated along with the other attributes of each trial. When defending his analysis among a myriad of critical tweets, Michael Walsh might have put it best: "The fragility index is not so much a statistical tool as a communication tool. We would all welcome alternative or additional tools that better communicate uncertainty".[5]

Priority should be given to the development of a version of the fragility index that is more applicable to time-to-event outcomes in clinical trials. We commend the developments reported by Bomze and Meirson and by Desnoyers and colleagues, and await publication of the operating characteristics of these methods.

We declare no competing interests.

*Joseph C Del Paggio, *Ian F Tannock*
ian.tannock@uhn.ca

Department of Medical Oncology, Thunder Bay Regional Health Sciences Centre and Northern Ontario School of Medicine, Thunder Bay, ON, Canada (JCDP); and Division of Medical Oncology, Princess Margaret Cancer Centre and University of Toronto, Toronto, ON M5G 2M9, Canada (IFT)

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 Kantarjian HM, DeAngelo DJ, Stelljes M, et al. Inotuzumab ozogamicin versus standard therapy for acute lymphoblastic leukemia. *N Engl J Med* 2016; **375:** 740–53.
4 Vera-Badillo FE, Napoleone M, Krzyzanowska MK, et al. Bias in reporting of randomised clinical trials in oncology. *Eur J Cancer* 2016; **61:** 29–35.
5 @lastwalsh. April 18, 2019. https://twitter.com/lastwalsh/status/1118866707697635328 (accessed Sept 13, 2019).