The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a retrospective analysis

Joseph C Del Paggio, Ian F Tannock

Summary

Background The fragility index of trial results—ie, the minimum number of changes from non-events to events resulting in loss of statistical significance—can provide a measure of confidence that a positive effect reported in a randomised controlled trial is real. We aimed to calculate the fragility index of randomised controlled trials supporting US Food and Drug Administration (FDA)-approved anticancer drugs.

Methods This is a retrospective analysis of phase 3, randomised, controlled trials supporting anticancer drugs that were approved by the FDA between Jan 1, 2014, and Dec 31, 2018. Two-arm studies with 1:1 randomisation and significant positive results for a time-to-event outcome were eligible for the fragility index calculation, which involves the iterative addition of an event to the experimental group (defined as the group with the smaller number of events in positive trials) and concomitant subtraction of a non-event from that group, until positive significance (defined as $p<0·05$ by Fisher's exact test) is lost.

Findings We identified 36 phase 3 randomised controlled trials, of which 17 (47%) were included in the fragility index analysis. The median fragility index was 2 (IQR 0–27). The fragility index was 2 or less in nine (53%) of 17 trials; for these trials, the fragility index was 1% or less of the total sample size. In five (29%) of 17 trials, the number lost to follow-up was more than the fragility index.

Interpretation 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 not a measure of effect, the fragility index might provide an additional means of assessing the robustness of clinical trial data.

Funding None.

Lancet Oncol 2019; 20: 1065–69

Published Online July 8, 2019 http://dx.doi.org/10.1016/S1470-2045(19)30338-9

Department of Medical Oncology, Thunder Bay Regional Health Sciences Centre and Northern Ontario School of Medicine, Thunder Bay, ON, Canada (J C Del Paggio MD); **and Division of Medical Oncology, Princess Margaret Cancer Centre and University of Toronto, Toronto, ON, Canada** (Prof I F Tannock MD)

Correspondence to: Prof Ian F Tannock, Division of Medical Oncology, Princess Margaret Cancer Centre, Toronto, ON M5G 2X9, Canada **ian.tannock@uhn.ca**

Introduction

Reliance on p values for establishing significance of results in clinical trials is contentious.[1] Outcomes meeting the arbitrary threshold of a p value less than 0·05 might not be clinically relevant, particularly when the difference in outcome does not provide substantial clinical benefit according to European Society for Medical Oncology or American Society of Clinical Oncology value scales.[2,3]

The statistical fragility of the results of randomised controlled trials can be represented by the ease with which its threshold p value shifts from significant ($p<0·05$) to non-significant ($p\geq0·05$) when experimental outcomes change from non-events to events. Walsh and colleagues[4] defined the fragility index as the minimum number of such changes, and the fragility index provides a measure of confidence that a positive effect reported in a randomised controlled trial comparing an experimental to a control treatment is real. The purpose of this study is to assess the fragility of phase 3 trials supporting recent US Food and Drug Administration (FDA)-approved anticancer drugs.

Methods

Study design

We reviewed FDA approvals for anticancer medicines between Jan 1, 2014 and Dec 31, 2018, publicly available at the FDA website. Using Google Scholar, we identified the phase 3 trials supporting each drug for the indication for which it was approved. For the fragility index analysis, we included only two-arm studies with 1:1 randomisation that reported significant positive primary outcome results for a time-to-event outcome[4] for the intention-to-treat population; secondary endpoints were assessed in cases in which the primary endpoint was not significant (ie, $p\geq0·05$ or the upper limit of the CI crossed 1). We abstracted information on trial design, observed numbers of events for the control and experimental groups for primary or secondary time-to-event outcomes, and the number of patients lost to follow-up. Data not available in the primary publication or its appendix were augmented by data in ClinicalTrials.gov or in Statistical Review and Evaluation documents on the FDA website.

The fragility index was calculated from a two by two contingency table by the iterative addition of an event to the experimental group (defined as the group with the smaller number of events in positive trials) and concomitant subtraction of a non-event from that same group, thereby maintaining a constant total number of events plus non-events, until positive significance (defined as $p<0·05$) was lost. p values were calculated

For the **FDA website**, see https://www.fda.gov

For **ClinicalTrials.gov**, see https://clinicaltrials.gov/

Research in context

Evidence before this study

We searched Google Scholar with the search terms "fragility index" and "oncology" or "FDA" for estimates of trial fragility in anticancer trials. Studies in areas of medicine other than oncology have shown that a change in a small number of events from negative to positive in the experimental group of a randomised controlled trial can lead to loss of significance; the number of events to provide loss of significance has been termed the fragility index. We found no previous estimates of trial fragility for randomised trials assessing anticancer drugs.

Added value of this study

We did a review of the FDA website to identify anticancer drugs approved in the 5-year interval of 2014–18. We identified two-arm phase 3 trials reporting significant positive results that supported FDA approval of an anticancer drug. We calculated the fragility index for each of these trials. To our knowledge, this is the first assessment of fragility index for randomised controlled trials in oncology.

Implications of all the available evidence

We found that many phase 3, randomised, controlled trials supporting FDA-approved anticancer drugs have a low fragility index, often less than 1% of the sample size or less than the number of patients lost to follow-up. Our results show that approval of many anticancer drugs is based on fragile evidence. The fragility index might provide an additional means of assessing the robustness of clinical trial data.

For the **online fragility index calculator** see https://clincalc.com/Stats/FragilityIndex.aspx

with Fisher's Exact Test.[4] An online fragility index calculator is available.

Role of the funding source

There was no funding source for this study. Both authors had full access to all data used in the study. The corresponding author had final responsibility for the decision to submit for publication.

Results

Between Jan 1, 2014, and Dec 31, 2018, the FDA approved 55 anticancer drugs for novel indications, of which 32 (58%) were for solid tumours. 29 (53%) of the drugs were approved on the basis of phase 1 or phase 2 trials. Five (9%) of the 55 approvals were based on the endpoint of overall survival, 19 (35%) were based on an alternative time-to-event endpoint, most often progression-free survival, 30 (55%) were based on measures of tumour response, and one (2%) was based on pharmacokinetic data. Phase 3 trials for the approved settings or indications were identified for 36 (65%) of 55 drugs.

17 (47%) of 36 phase 3 randomised controlled trials met the inclusion criteria for fragility index analysis (table); we could not calculate the fragility index for the remaining 19 phase 3 trials because of unequal allocation between groups (16 trials) or a statistically negative time-to-event endpoint (three trials). The median sample size for the 17 eligible randomised controlled trials was 452 (range 220–2840). The primary endpoint was used for fragility index analysis in 16 of the 17 eligible trials;[5–16,18–21] for the remaining trial,[17] the fragility index was calculated using the secondary progression-free survival endpoint, because the primary overall survival endpoint was not significant (table).

The median fragility index for the 17 studies was 2 (IQR 0–27)—ie, a median of two events was required to change the results of the endpoint analysis from significant to non-significant (figure). Fragility index was 2 or less in nine (53%) of 17 trials;[5,8–11,15,17,19,20] for these trials, the fragility index was 1% or less of the total sample size. For the six trials with a fragility index of 0 (ie, Fisher's exact test p>0·05), the χ^2 test (one trial[9]) and stratified log-rank test (five trials[5,10,15,17,19]) had been used to calculate the reported significant p value. In six (35%) of 17 trials,[9,10,12,16,17,21] the number of patients lost to follow-up was two or more (median 1 [IQR 0–2; range 0–68]); the number lost to follow-up was more than the respective fragility index in five (29%) of the 17 trials.[9,10,16,17,19] Of the 17 drugs tested in the eligible trials included in the fragility index analysis, only one drug (daratumumab) was supported by more than one positive phase 3 trial in that setting and indication (table).[13,22]

Discussion

In our retrospective analysis, we show that about half of the phase 3 trials supporting FDA-approved anticancer drugs have a low fragility index and are vulnerable to losing significance with a change in designation of very few events, often a change in event number less than 1% of the respective trial sample size. The change in number of events required for fragility is also often smaller than the number of patients lost to follow-up, raising concerns about a statistical change in the results had these patients been assessed to their endpoints.

To our knowledge, no previous studies have estimated the fragility index for oncology trials or for trials strictly supporting FDA-approved medications. The fragility index has been applied to other randomised controlled trials, including those assessing spinal surgery,[23] critical care,[24] and heart failure,[25] and to trials supporting clinical practice guidelines.[26,27] These studies are consistent in showing that many randomised controlled trials are fragile, and several investigators have recommended adoption of the fragility index in reporting clinical trial outcomes.[24,27,28] Trials with large fragility indexes are present in our cohort; however, most trials were powered to detect differences in progression-free survival, which is subject to biases of clinical and radiological assessment, as well as informative

	Approval date	Disease site or indication	Trial	Endpoint	Experimental sample size	Experimental event number*	Control sample size	Control event number*	p value†	Fragility index
Ceritinib	April 2014	Non-small-cell lung cancer	ASCEND-5[5]	Progression-free survival	115	83	116	89	<0·0001	0
Idelalisib	July 2014	Chronic lymphocytic leukaemia	GS-US-312-0116[6]	Progression-free survival	110	12	110	53	<0·001	26
Panobinostat	February 2015	Multiple myeloma	PANORAMA1[7]	Progression-free survival	387	207	381	260	<0·0001	31
Dinutuximab	March 2015	Neuroblastoma	ANBL0032[8]	Event-free survival	113	33	113	50	0·0115	2
Elotuzumab	November 2015	Multiple myeloma	ELOQUENT-2[9]	Progression-free survival	321	179	325	205	<0·001	0
Necitumumab	November 2015	Squamous non-small-cell lung cancer	SQUIRE[10]	Overall survival	545	418	548	442	0·012	0
Ixazomib	November 2015	Multiple myeloma	TOURMALINE-MM1[11]	Progression-free survival	360	129	362	157	0·012	2
Cobimetinib	November 2015	Melanoma	coBRIM[12]	Progression-free survival	247	79	248	128	<0·001	27
Daratumumab	November 2015	Multiple myeloma	CASTOR‡[13]	Progression-free survival	251	67	247	122	<0·0001	35
Venetoclax	April 2016	Chronic lymphocytic leukaemia	MURANO[14]	Progression-free survival	194	32	195	114	<0·001	62
Midostaurin	April 2017	Acute myeloid leukaemia	RATIFY[15]	Overall survival	360	171	357	186	0·002	0
Neratinib	July 2017	Breast cancer	ExteNET[16]	Invasive disease-free survival	1420	70	1420	109	0·0091	12
Inotuzumab ozogamicin	August 2017	Acute lymphocytic leukaemia	INO-VATE ALL[17]	Progression-free survival§	164	129	162	128	<0·001	0
Lutetium (^{177}Lu) oxodotreotide	January 2018	Gastroenteropancreatic neuroendocrine tumours	NETTER-1[18]	Progression-free survival	116	23	113	68	0·004	32
Mogamulizumab	August 2018	Mycosis fungoides or Sézary syndrome	MAVORIC[19]	Progression-free survival	186	110	186	122	<0·0007	0
Duvelisib	September 2018	Chronic lymphocytic leukaemia	DUO[20]	Progression-free survival	160	93	159	110	<0·0001	1
Dacomitinib	September 2018	Non-small-cell lung cancer	ARCHER 1050[21]	Progression-free survival	227	136	225	179	<0·0001	27

FDA=US Food and Drug Administration. *Event numbers established by independent review committees were abstracted preferentially when available. †Calculated by the statistical methods in each trial. ‡A second phase 3 trial, POLLUX,[22] was published in the same treatment setting; it has a progression-free survival fragility index of 42. §Significant secondary endpoint.

***Table*: Fragility index calculated for 17 phase 3 trials with 1:1 randomisation supporting drugs approved by the FDA between 2014 and 2018**

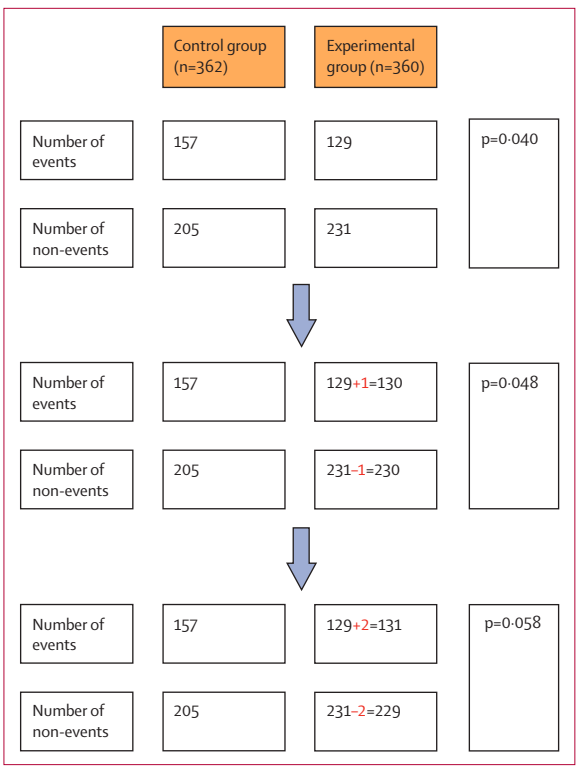

Figure: **Example of fragility index calculation for the phase 3 trial TOURMALINE-MM1[11]**

p values are calculated by Fisher's exact test, whereas the p value in the original study was calculated as 0·012 using the stratified log-rank test. The fragility index in this example is 2, which is the number of non-events required to convert to events so that the difference between the control and experimental groups no longer meets significance at the α=0·05 level using Fisher's exact test.

censoring, where there is loss of patients to follow-up before meeting criteria of progression.[29] A low fragility index in trials using progression-free survival or similar endpoints might be one of several factors leading to poor correlation with overall survival.

In principle, the p value is an indication of the compatibility between data from a trial and the prespecified statistical model: smaller p values imply greater statistical incompatibility of the data with the null hypothesis—a postulate of no difference between outcomes of the experimental and control group.[30] The p value depends on assumptions. The log-rank test used in survival analysis has the advantage that it accounts for events over time, but it relies on the assumption that the hazard ratio of two treatments is constant over time (ie, proportional hazards). Fisher's exact test (used to calculate the fragility index) has the disadvantage that it does not account for the time at which events occurred,[31] but it does not require proportional hazards, a condition that is not satisfied, for example, when survival curves cross. Fragility index calculations of zero (indicting a p≥0·05 by Fisher's exact test) were possible for trials reporting significance based on log-rank and other tests: these indicate extreme fragility.

The approval by the FDA of anticancer drugs considers the totality of evidence relating to their effectiveness in the context of the illness for which the drug is intended, the risks of the drug and management thereof, the uncertainties in extrapolating clinical data to the real world, and the applicable laws and regulations. Clinical data from phase 2 trials as well as phase 3, randomised, controlled trials might be analysed in the approval process; however, results of phase 2 trials can be misleading, and phase 3, randomised, controlled trials are regarded as providing the highest level of evidence relating to clinical benefit. For the drugs approved in the 5-year period under analysis in this study, it was rare that approval was supported by more than one phase 3, randomised, controlled trial.

This study is limited by its small sample size, necessitated by the number of oncological drugs approved by the FDA within the study period, as well as the 1:1 randomisation required for the fragility index calculation.[4] The operating characteristics of the fragility index also limit its use in time-to-event data: in situations where the number of events is similar between two groups, but a difference in timing exists, the fragility index might be overly sensitive in concluding fragility.[4] Finally, since a strong relationship exists between the p value and the fragility index,[32] caution must be taken in concluding the robustness of a clinical trial on fragility index alone without a broader context (eg, statistical design, effect size, CIs, and minimal important differences). As exemplified by the outcome data we presented, extreme fragility can be noted in situations in which the absolute difference in numbers of events between the experimental and control groups is quite large—a difference that might be considered clinically meaningful. In general, however, larger overall sample sizes increase the fragility index,[28] which speaks to the aforementioned correlation between the fragility index and p value.

The finding that many phase 3, randomised, controlled trials supporting FDA-approved anticancer drugs have a low fragility index challenges the confidence in concluding superiority for these drugs over control treatments. Many FDA-approved drugs have been shown to be of low clinical value,[33,34] and measuring the robustness of clinical trial data to support their high cost is paramount. The fragility index, like the p value, should not be interpreted as a measure of effect, but it can shed some light on the strength of statistical conclusions.

Contributors

All authors contributed to the design of the study. JCD acquired the data for the study, performed the analyses, and drafted the report. All authors interpreted the results and contributed to the writing of the final report.

Declaration of interests

We declare no competing interests.

References

1 Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019; **567:** 305–07.

2 Cherny NI, Dafni U, Bogaerts J, et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Ann Oncol* 2017; **28:** 2340–66.

3 Schnipper LE, Davidson NE, Wollins DS, et al. Updating the American Society of Clinical Oncology value framework: revisions and reflections in response to comments received. *J Clin Oncol* 2016; **34:** 2925–34.

4 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.

5 Shaw AT, Kim TM, Crinò L, et al. Ceritinib versus chemotherapy in patients with ALK-rearranged non-small-cell lung cancer previously given chemotherapy and crizotinib (ASCEND-5): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* 2017; **18:** 874–86.

6 Furman RR, Sharman JP, Coutre SE, et al. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N Engl J Med* 2014; **370:** 997–1007.

7 San-Miguel JF, Hungria VT, Yoon SS, et al. Panobinostat plus bortezomib and dexamethasone versus placebo plus bortezomib and dexamethasone in patients with relapsed or relapsed and refractory multiple myeloma: a multicentre, randomised, double-blind phase 3 trial. *Lancet Oncol* 2014; **15:** 1195–206.

8 Yu AL, Gilman AL, Ozkaynak MF, et al. Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. *N Engl J Med* 2010; **363:** 1324–34.

9 Lonial S, Dimopoulos M, Palumbo A, et al. Elotuzumab therapy for relapsed or refractory multiple myeloma. *N Engl J Med* 2015; **373:** 621–31.

10 Thatcher N, Hirsch FR, Luft AV, et al. Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controlled phase 3 trial. *Lancet Oncol* 2015; **16:** 763–74.

11 Moreau P, Masszi T, Grzasko N, et al. Oral ixazomib, lenalidomide, and dexamethasone for multiple myeloma. *N Engl J Med* 2016; **374:** 1621–34.

12 Ascierto PA, McArthur GA, Dréno B, et al. Cobimetinib combined with vemurafenib in advanced BRAF(V600)-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. *Lancet Oncol* 2016; **17:** 1248–60.

13 Palumbo A, Chanan-Khan A, Weisel K, et al. Daratumumab, bortezomib, and dexamethasone for multiple myeloma. *N Engl J Med* 2016; **375:** 754–66.

14 Seymour JF, Kipps TJ, Eichhorst B, et al. Venetoclax-rituximab in relapsed or refractory chronic lymphocytic leukemia. *N Engl J Med* 2018; **378:** 1107–20.

15 Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N Engl J Med* 2017; **377:** 454–64.

16 Chan A, Delaloge S, Holmes FA, et al. Neratinib after trastuzumab-based adjuvant therapy in patients with HER2-positive breast cancer (ExteNET): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* 2016; **17:** 367–77.

17 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.

18 Strosberg J, El-Haddad G, Wolin E, et al. Phase 3 trial of ^{177}Lu-dotatate for midgut neuroendocrine tumors. *N Engl J Med* 2017; **376:** 125–35.

19 Kim YH, Bagot M, Pinter-Brown L, et al. Mogamulizumab versus vorinostat in previously treated cutaneous T-cell lymphoma (MAVORIC): an international, open-label, randomised, controlled phase 3 trial. *Lancet Oncol* 2018; **19:** 1192–04.

20 Flinn IW, Hillmen P, Montillo M, et al. The phase 3 DUO trial: duvelisib vs ofatumumab in relapsed and refractory CLL/SLL. *Blood* 2018; **132:** 2446–55.

21 Wu YL, Cheng Y, Zhou X, et al. Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): a randomised, open-label, phase 3 trial. *Lancet Oncol* 2017; **18:** 1454–66.

22 Dimopoulos MA, Oriol A, Nahi H, et al. Daratumumab, lenalidomide, and dexamethasone for multiple myeloma. *N Engl J Med* 2016; **375:** 1319–31.

23 Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. *Spine J* 2015; **15:** 2188–97.

24 Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The fragility index in multicenter randomized controlled critical care trials. *Crit Care Med* 2016; **44:** 1278–84.

25 Docherty KF, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. How robust are clinical trials in heart failure? *Eur Heart J* 2017; **38:** 338–45.

26 Chase Kruse B, Matt Vassar B. Unbreakable? An analysis of the fragility of randomized trials that support diabetes treatment guidelines. *Diabetes Res Clin Pract* 2017; **134:** 91–105.

27 Edwards E, Wayant C, Besas J, Chronister J, Vassar M. How fragile are clinical trial outcomes that support the CHEST clinical practice guidelines for VTE? *Chest* 2018; **154:** 512–20.

28 Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg* 2019; **154:** 74–79.

29 Tannock IF, Amir E, Booth CM, et al. Relevance of randomised controlled trials in oncology. *Lancet Oncol* 2016; **17:** e560–67.

30 Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat* 2016; **70:** 129–33.

31 Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. *Crit Care* 2004; **8:** 389–94.

32 Carter RE, McKie PM, Storlie CB. The fragility index: a p-value in sheep's clothing? *Eur Heart J* 2017; **38:** 346–48.

33 Booth CM, Del Paggio JC. Approvals in 2016: questioning the clinical benefit of anticancer therapies. *Nat Rev Clin Oncol* 2017; **14:** 135–36.

34 Tibau A, Molto C, Ocana A, et al. Magnitude of clinical benefit of cancer drugs approved by the US Food and Drug Administration. *J Natl Cancer Inst* 2018; **110:** 486–92.