Frequent fragility of randomized controlled trials for HCC treatment

Zhang Hao
Guangzhou medical University
Li Jingtao
Shaanxi University of Chineses University
Wenting Zeng (13609645855@163.com)
Guangzhou Medical College First Affiliated Hospital https://orcid.org/0000-0003-2372-5066

Research article

Keywords: Fragility index, Randomized controlled trials, Time-to-event outcome

Posted Date: January 9th, 2020

DOI: https://doi.org/10.21203/rs.2.20418/v1

Version of Record: A version of this preprint was published at BMC Cancer on April 9th, 2021. See the published version at https://doi.org/10.1186/s12885-021-08133-8.

Abstract

Background

The fragility index (FI) of trial results can provide a measure of confidence in the positive effects reported in randomized controlled trials (RCTs). The aim of this study was to calculate the FI of RCTs supporting HCC treatments.

Methods

A methodological systematic review of RCTs in HCC treatments was conducted. Two-arm studies with randomized and positive results for a time-to-event outcome were eligible for the FI calculation.

Results

A total of 11 trails were included in this analysis. The median FI was 0 (range 0-19). FI was ≤ 5 in 8 (72.73%) of 11 trials; in those trials the fragility quotient was ≤ 1%.

Conclusion

Many phase 3 RCTs supporting HCC treatments have a low FI, which challenges the confidence in concluding the superiority of these drugs over control treatments.

Background:

Modern medicine is built on evidence-based clinical practice, with randomized controlled trials (RCTs) forming the foundation of such evidence. Because RCTs play important roles in governing clinical practice, the robustness of their results is critical. The results of clinical trials must be valid, reproducible, and repeatable; however, in the context of clinical research, reproducibility and replicability are generally under-researched topics. Historically, P values have been used to indicate statistical the significance of results in clinical trials[1]. Nevertheless, this approach has some significant limitations and has been heavily criticized for being simplistic, with frequent misapplication and misinterpretation[2].

The fragility index (FI) is a novel tool, which was developed to assess the robustness of statistically significant dichotomous outcomes from RCTs[3]. It is defined as the minimum number of patients receiving experimental treatment whose status would have to change from a non-event to an event to nullify a meaningful result. A higher FI represents a relativiely robust outcome and indicates that the statistical significance of a given outcome hinges on a greater number of events, whereas a lower FI indicates that the statistical significance of a given outcome depends on only a few events, which suggests a more fragile outcome.

The recommendation of new drugs or treatments for use in clinical practice mainly depends on the results of phase 3 clinical trials. Thus, this study performed a retrospective analysis to assess the wider implications of the FI in the findings of HCC treatments in phase 3 clinical trials.

Methods:

This study conducted a methodological systematic review of phase 3 RCTs for HCC treatment. The search terms used were (hepatocarcinoma OR "liver cancer" OR HCC) AND ("phase 3" OR “phase III”). Only articles published in English were searched for using PubMed search engine and Medline database until August 1, 2019. This analysis is reported according to the Preferred Reporting Items for Systematic Reviewers and Meta-Analyses guidelines (Supplementary table) [4].

For the FI analysis, only two-arm studies with randomization that reported significant positive results with primary or secondary outcomes were included. Data was obtained on trial design, trial number, and the observed numbers of events for the control and experimental groups in primary or secondary time-to-event outcomes. Any data that was unavailable in the publication or its appendix was augmented by data form the ClinicalTrials.gov. The FI was calculated from a two-by-two contingency table by the iterative addition of an event to the experimental group, which was determined using a web-based fragility calculator (available at http://www.clincalc.com/Stats/FragilityIndex.aspx). P values were calculated using Fisher’s Exact Test. A sample of FI is presented in Fig. 1.

The fragility quotient (FQ) is a metric, that accounts for the FI in the context of sample size [5]. It is described as the FI divided by the total sample size. The usefulness of the FQ lies in its ability to allocate an objective value to the results of subjective importance, and it may be assigned to an outcome with a given FI in a certain sample size. In other words, the FQ assesses the robustness of the FI.

Results:

This study identified 114 records through a series of PubMed searches (Fig. 2). After an initial screening of abstracts and a full-text review of the studies, 11 articles were included in the fragility analysis (Table 1). The median sample size for the 11 eligible RCTs was 395 (range 205–707), and the median FI for the 11 studies was 0 (range 0–19). The FI ≤ was 5 in 8 (72.73%) of 11 trials[6 – 13], and those trials had FQ ≤ 1%.

Table 1
Fragility index calculated for 11 phase 3 trials with 1:1 randomization for HCC treatment.

Author	Study name	Clinical Trial	Experimental Treatment vs. Control	Endpoint	Experimental sample size	Experimental event number	Control sample size	Control event number	P vaule	Fragility index
Zhu AX et al[9].	REACH-2	NCT02435433	Ramucirumab vs. Placebo	Primary endpoint: Overall survival	197	142	95	74	0.0199	0
Abou-Alfa GK et al[7].	CELESTIAL	NCT01908426	Cabozantinib vs. Placebo	Primary endpoint: Overall survival	470	317	237	167	0.0049	0
Kudo M et al [15].	SILIUS	NCT01214343	Sorafenib plus HAIC (hepatic arterial infusion chemotherapy) vs. Sorafenib	Primary outcome: Overall response	102	37	103	18	0.003	7
Wang Z et al [13].	NA	NCT01966133	adjuvant TACE vs. No adjuvant TACE	Primary endpoint: Recurrence-free survival	140	46	140	82	0.01	19
Bruix J et al [11].	RESORCE	NCT01774344	Regorafenib vs. Placebo	Primary endpoint: Overall survival	379	146	194	54	0.00002	5
Lee JH et al [5].	NA	NCT00699816	CIK cell agent vs. No CIK cell agent	Primary end point: Recurrence-free survival	114	69	112	59	0.01	0
Llovet JM et al [12].	SHARP	NCT00105443	Sorafenib vs. Placebo	Primary endpoint: Overall survival	299	44	303	33	0.00583	0
Wei W et al [6].	NA	NCT02788526	Hepatectomy plus TACE vs. Hepatectomy	Primary endpoint: Disease-free survival	116	83	118	85	0.02	0
Geissler EK et al [10].	NA	NCT00355862.	Liver transplantation with sirolimus vs. Liver transplantation	Secondary endpoint: Overall survival	252	242	256	234		1
Llovet JM et al [14].	BRISK-PS	NCT00825955	Sorafenib plus Brivanib vs. Sorafenib plus Placebo	Secondary outcome: Disease-control rate	263	160	132	52	< 0.001	15
Zhu AX et al [8].	EVOLVE-1	NCT01035229	Sorafenib plus Everolimus vs. sorafenib plus Placebo	Secondary outcome: Disease-control rate	362	203	184	83	0.01	4

Eight studies in the fragility analysis were for primary outcome results. Five (40%) had primary outcome trials with a FI of 0 (Fisher's exact test $p > 0.05$), for which a stratified log-rank test was used to calculate the reported significant P value[6 – 8, 10, 13], and six (75%) trials had an FQ < 1%[6–8, 10, 12, 13]. The article with the highest FI fragility index of 19 was published in the Clinical Cancer Research[14]. However, this study was not a multiple center trial. The remaining 3 studies were evaluated with inferior outcome results, whereas non significant differences were found in the primary outcome results. The studies of the FI were 1, 4, and 15, respectively[11, 15, 16], and of which two (66.67%) had an FQ < 1%[11, 16].

Discussion:

To the best of our knowledge, FI investigation for HCC trials has not been performed. The FI has been evaluated in other RCTs, such as emergency medicine[17], giant cell arteritis[18], and Clinical Practice Guidelines[19]. These studies consistently show that many RCTs are fragile, and several researchers have recommended that FI should be adopted in reporting clinical trial outcomes[18, 19], our study showed that most results from the randomized trials were far more fragile.

This retrospective analysis demonstarated that over 70% of the phase 3 trials supporting HCC treatments had a low FI; however, they are vulnerable to losing their significance with just a small change in the designation of a small number of events, often equating to < 1% of the sample size in an experimental group. As clinical practices or the use of drugs approved by Food and Drug Administration are developed on the results of phase 3 clinical trials, the change in the number of events required for fragility raises concerns about a statistical change in the results.

RCTs, particularly phase 3 clinical trials, are likely to remain an important evidence base for clinicians' practice. Despite this, the statistical methodology used to establish significance in such clinical trials has barely evolved. In principle, the P value is an indication of the compatibility among data from a trial; a smaller P value implies a greater statistical incompatibility of the result with the null hypothesis (an estimation of no difference between the experimental and control group[20]). However, this approach has been greatly criticized for being simplistic, and has frequently been misinterpreted[21]. The log-rank test used in survival data analysis has advantage in that it accounts for events, but it relies on the assumption that the hazard ratio of two treatments remains constant over time. Fisher's exact test, which is used to calculate the FI, has the disadvantage of not accounting for the time-to-event[22]. Thus, the FI is simplistic in its application and resolves some of these shortcomings.

Although the FI and FQ do provide a relative wealth of information when consider alongside other metrics, this study again emphasizes the limitations of the FI itself. First, clinical trials must obtain significant in effects in the treatment group, which could to be analyzed by the FI. Many non-inferiority studies cannot be included in this analysis, such as the E7080 trials of lenvatinib for HCC, which produced the same treatment results as sorafenib[23]. Second, because the FI relies on P value, it is essentially an extension of the most frequent approach to data analysis. Thus, it cannot be applied to an outcome of a continuous variable. Third, although many time-to-event outcomes are usually dichotomous, such as mortality, and survival, etc, the FI does not account for the difference in outcomes over time. Particularly in longer studies with variable follow-up time periods, analyses that account for time (such as a Kaplan–Meier curve, or a Cox proportional hazards model) are more appropriate than a simple binary outcome analysis. Finally, there is no specific cut-off value or lower limit of the FI to classify a study as "either fragile" or "robust".

Conclusion:

The outcomes of many phase 3, RCTs supporting HCC treatments with a low FI challenges the confidence in concluding the superiority of these drugs over control treatments.

Abbreviations:

HCC
hepatocarcinoma
RCTs
randomized controlled trials
FI
fragility index
FQ
fragility quotient

Declarations:

Ethics approval and consent to participate Not applicable

Availability of data and materials Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests The authors declare that they have no competing interests

Funding This study was partly supported by the grant from the National Natural Science Foundation of China (No. 81603612); the Science and Technology Department of Shaanxi Province (NO.2018KJXX-093), and Innovation Team of Shaanxi University of Traditional Chinese Medicine (NO.2019-YL05). The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions All authors were involved in the study design, including setting up the keywords search and project protocol. ZH and LJT collected the data information. ZH draft manuscript. ZWT and Li Jingtao were responsible for the supervision of the project and revise of the manuscript. All authors were finally approval of the manuscript.

Acknowledgements Not applicable

Consent for Publication Not applicable

References:

1. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature 2019;567:305-307.
2. Ioannidis J. The Proposal to Lower P Value Thresholds to .005. JAMA 2018;319:1429-1430.

3. 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;67:622-628.
4. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLOS MED 2009;6(7):e1000097.
5. Ahmed W, Fowler RA, McCredie VA. Does Sample Size Matter When Interpreting the Fragility Index? Crit Care Med 2016;44:e1142-e1143.
6. Lee JH, Lee JH, Lim YS, et al. Adjuvant immunotherapy with autologous cytokine-induced killer cells for hepatocellular carcinoma. Gastroenterology 2015;148:1383-1391.
7. Wei W, Jian PE, Li SH, et al. Adjuvant transcatheter arterial chemoembolization after curative resection for hepatocellular carcinoma patients with solitary tumor and microvascular invasion: a randomized clinical trial of efficacy and safety. Cancer Commun (Lond) 2018;38:61.
8. Abou-Alfa GK, Meyer T, Cheng AL, et al. Cabozantinib in Patients with Advanced and Progressing Hepatocellular Carcinoma. N Engl J Med 2018;379:54-63.
9. Zhu AX, Kudo M, Assenat E, et al. Effect of everolimus on survival in advanced hepatocellular carcinoma after failure of sorafenib: the EVOLVE-1 randomized clinical trial. JAMA 2014;312:57-67.
10. Zhu AX, Kang YK, Yen CJ, et al. Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased alpha-fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 2019;20:282-296.
11. Geissler EK, Schnitzbauer AA, Zulke C, et al. Sirolimus Use in Liver Transplant Recipients With Hepatocellular Carcinoma: A Randomized, Multicenter, Open-Label Phase 3 Trial. Transplantation 2016;100:116-125.
12. Bruix J, Qin S, Merle P, et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;389:56-66.
13. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378-390.
14. Wang Z, Ren Z, Chen Y, et al. Adjuvant Transarterial Chemoembolization for HBV-Related Hepatocellular Carcinoma After Resection: A Randomized Controlled Study. Clin Cancer Res 2018;24:2074-2081.
15. Llovet JM, Decaens T, Raoul JL, et al. Brivanib in patients with advanced hepatocellular carcinoma who were intolerant to sorafenib or for whom sorafenib failed: results from the randomized phase III BRISK-PS study. J Clin Oncol 2013;31:3509-3516.
16. Kudo M, Ueshima K, Yokosuka O, et al. Sorafenib plus low-dose cisplatin and fluorouracil hepatic arterial infusion chemotherapy versus sorafenib alone in patients with advanced hepatocellular carcinoma (SILIUS): a randomised, open label, phase 3 trial. Lancet Gastroenterol Hepatol. 2018;3(6):424-432..
17. Brown J, Lane A, Cooper C, et al. The Results of Randomized Controlled Trials in Emergency Medicine Are Frequently Fragile. Ann Emerg Med 2019;73:565-576.
18. Berti A, Cornec D, Medina IJ, et al. Treatments for giant cell arteritis: Meta-analysis and assessment of estimates reliability using the fragility index. Semin Arthritis Rheum 2018;48:77-82.
19. Edwards E, Wayant C, Besas J, et al. How Fragile Are Clinical Trial Outcomes That Support the CHEST Clinical Practice Guidelines for VTE? Chest 2018;154:512-520.
20. Demidenko E. The p-Value You Can't Buy. Am Stat 2016;70:33-38.
21. Sterne JA, Davey SG. Sifting the evidence-what's wrong with significance tests? BMJ 2001;322:226-231.
22. Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. Crit Care 2004;8:389-394.
23. Kudo M, Finn RS, Qin S, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. Lancet. 2018;391(10126):1163-1173.

Figures

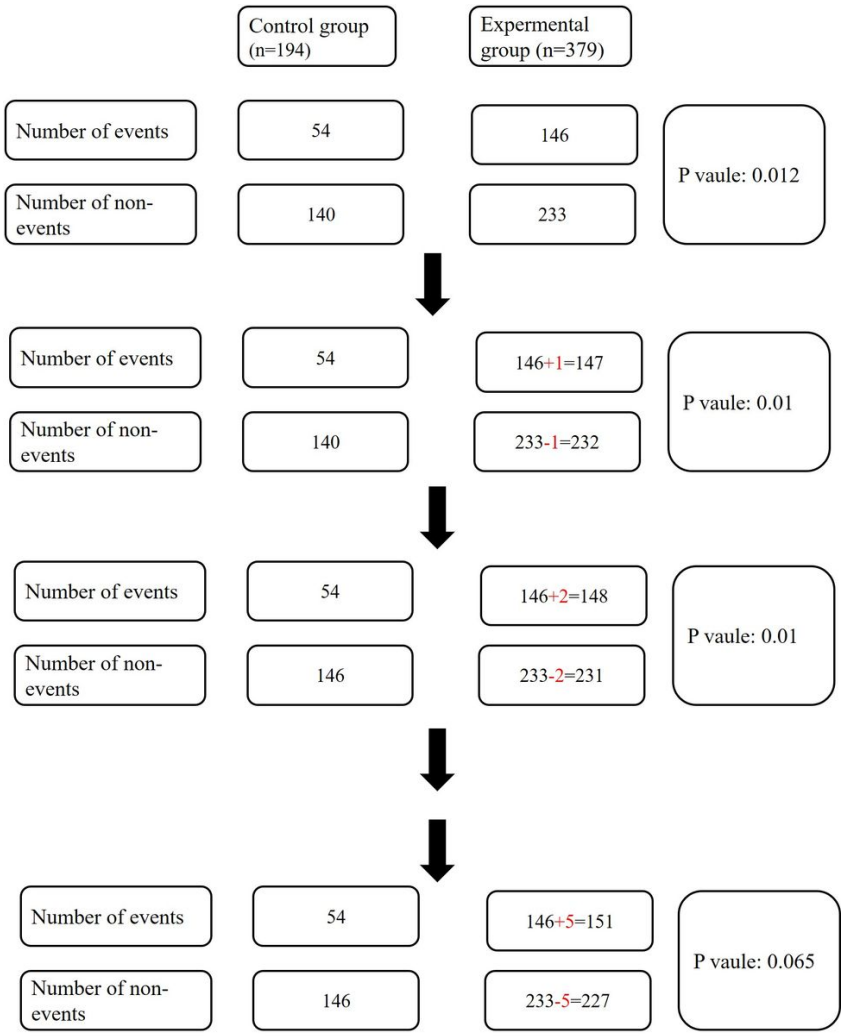

Figure 1

Example of fragility index calculation for the phase 3 trial RESORCE reported by Bruix J, et al (11).

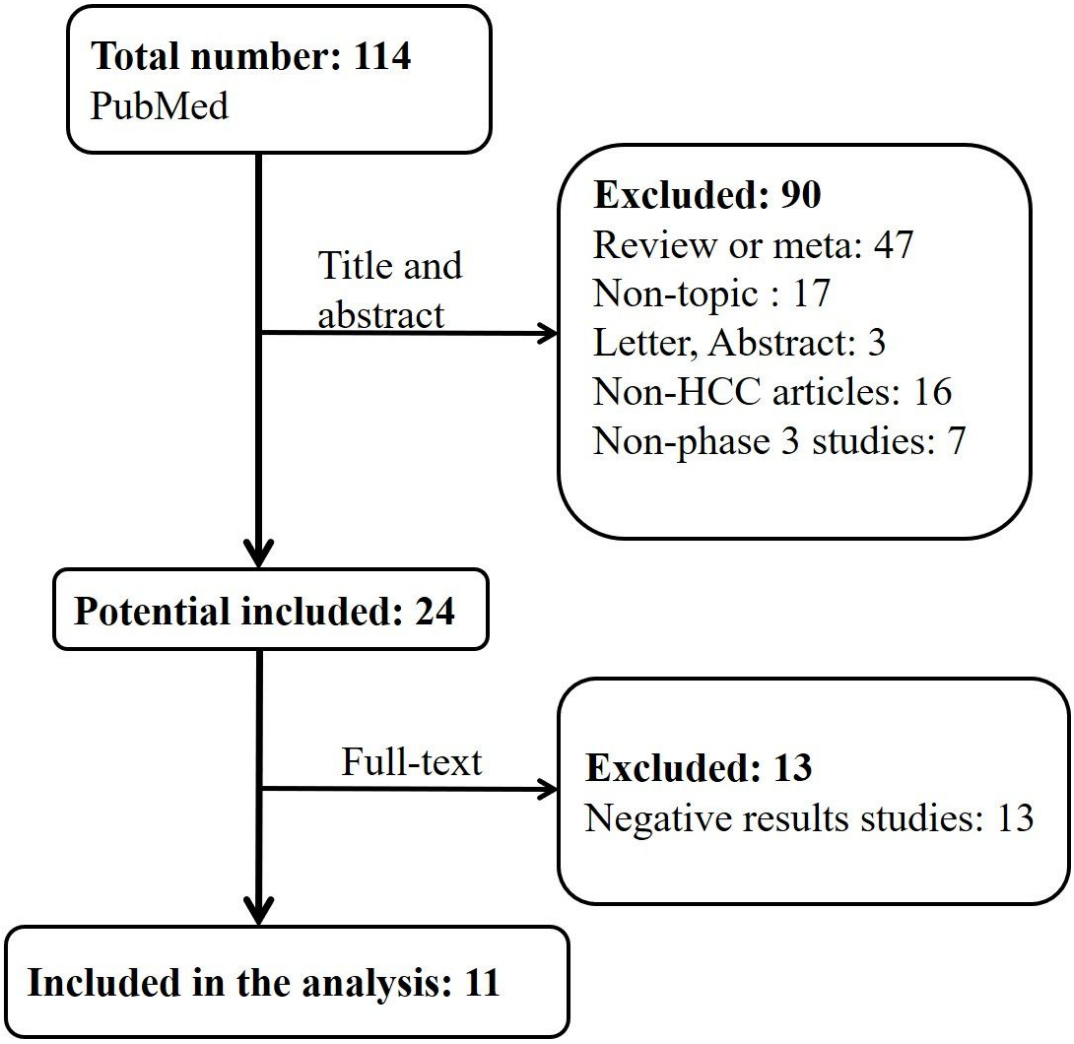

Figure 2

Flow chart for included studies.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuprementaryTabelPRISMA2009checklist.doc