Assessment of robustness of randomized controlled trials for the treatment of cholangiocarcinoma in three domains: survival-inferred fragility index, restricted mean survival time, and the spin effect

Nir Horesh, David Bomze, Chetana Lim, Gal Markel, Tomer Meirson, Daniel Azoulay

HAL Id: hal-03525790

https://hal.sorbonne-universite.fr/hal-03525790v1

Submitted on 14 Jan 2022

Assessment of robustness of randomized controlled trials for the treatment of cholangiocarcinoma in three domains: survival-inferred fragility index, restricted mean survival time, and the spin effect

Nir Horesh[1,2], David Bomze[2], Chetana Lim[3], Gal Markel[2,4], Tomer Meirson[4,5,6], Daniel Azoulay[1,7]

[1]Unit of Hepatobiliary and Pancreatic Surgery and Transplantation, Sheba Medical Center, Faculty of Medicine Tel Aviv University, Tel Aviv, Israel; [2]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [3]Centre Hépato-Biliaire, Department of Hepatobiliary and Pancreatic Surgery and Transplantation, Pitié Salpetriere Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France; [4]Davidoff Cancer Center, Rabin Medical Center-Beilinson Hospital, Petah Tikva, 49100, Israel; [5]Shamir Medical Center (Assaf Harofeh), Rishon Lezion, Israel; [6]The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel; [7]Centre Hépato-Biliaire, Department of Hepatobiliary and Pancreatic Surgery and Transplantation, Paul Brousse Hospital, Assistance Publique-Hôpitaux de Paris, Université Saclay, Villejuif, France

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: N Horesh, D Azoulay, D Bomze; (V) Data analysis and interpretation: D Bomze, T Meirson; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Daniel Azoulay. Centre Hépato-Biliaire, Hôpital Paul Brousse, Assistance Publique-Hôpitaux de Paris, Université Saclay, Villejuif, France. Email: daniel.azoulay@aphp.fr.

Importance: The vast majority of patients with cholangiocarcinoma (CC) have advanced disease at diagnosis and are candidates for palliative treatment only.

Objective: The robustness of the randomized controlled trials regarding the treatment of CC are assessed.

Evidence Review: A systematic review of all randomized control trials (RCT) of treatments for both intra- and extrahepatic CC between 2010 and 2020 was performed. The survival-inferred fragility index (SIFI, the minimum number of reassignments of the best survivors between arms that would overturn the statistical outcomes) was calculated. In addition, the gain, or loss, in survival in RCTs was evaluated by the restricted mean survival time (RMST) difference. Finally, the level of spin i.e., misrepresentation of study outcomes, was measured in inconclusive studies to assess distorted reporting strategies.

Findings: Out of 6,167 studies retrieved, 11 could be retained for full text revision (7 with both intra- and extrahepatic CC, 3 with peri-hilar CC, and 1 with peri-hilar or distal CC). Only 3 studies included resected patients (2 with both intra- and extrahepatic CC and 1 with peri-hilar or distal CC). Nine studies investigated systemic chemotherapy (including 3 after surgical resection), one study evaluated photodynamic therapy, and another investigated the use of an endoscopically inserted stent in the biliary tract. The median SIFI was –2 (IQR –6.25, –0.25) across all studies. Overall, the median RMST difference was 0.56 months (IQR 0.10, 0.95). Finally, for inconclusive studies, the level of spin was high, moderate, and low in respectively 12.5%, 25%, and 62.5% of the studies.

Conclusions and Relevance for Reviews: RCTs of CC showed a low degree of robustness with a frequent proportion of associated spin.

Keywords: Cholangiocarcinoma (CC); randomized controlled trials; robustness of evidence; survival-inferred fragility index (SIFI); restricted mean survival time (RMST); spin effect

Introduction

Surgery is the only potential curative option for cholangiocarcinoma (CC) but is indicated in only ~25% of patients (1), whereas most patients are diagnosed at a deemed unresectable stage and candidates for palliative treatment only. The current guidelines, based on fairly small scale studies (2), recommend GEMOX and FOLFOX as first- and second-line chemotherapy, respectively (1). The evidence generated by the few available RCTs evaluating possible treatments for advanced CC that drives clinical decisions, relies on p values and hazard ratios (HR) despite the ongoing debate about the former (3-5) and the frequent variation with time of the latter, that might violate the Cox proportional hazards model assumption. To calculate the stability of the statistical calculation and evaluating the effect size more reliably, in particular that patients live longer, alternative measures of treatment outcomes have been developed. The latter include the survival-inferred fragility index (SIFI), the restricted mean survival time (RMST), and the spin. The SIFI, is the minimum number of reassignments of the best survivors between arms that would overturn the statistical outcomes (6-8). The RMST is a statistical calculation of the time to event corresponding to the area under the Kaplan-Meir survival curve limited to a specific time point. When the difference between the two curves in the RMST estimate is positive or negative, it can be viewed as a gain or loss of survival time, respectively. The measurement of the spin effect (3) found in inconclusive RCTs emphasizes on how the researchers, reporting results in a more favorable light, might propose misleading arguments that in turn mislead treatment decisions and hamper the development of new treatment strategies.

The purpose of this study was to evaluate the SIFI and the RMST of RCTs comparing novel treatments with control or standard treatments for CC and evaluate the level of spin in inconclusive RCTs. We present the following article in accordance with the PRISMA reporting checklist (available at https://dx.doi.org/10.21037/hbsn-21-118).

Methods

A systematic database search (Embase, Medline, Cochrane Library and Google Scholar) of the literature was performed using both free-text and Medical Subject Headings (MeSH) terms (cholangiocarcinoma and chemotherapy, surgery or radiotherapy, cholangiocarcinoma, perihilar cancer, Klatskin tumor) during the past decade (2010–2020). Both types of CC (i.e., intra- and extrahepatic) were included for evaluation because of the low number of available RCTs.

Articles were selected for inclusion in the analysis of the evidence based on the following criteria:

- ❖ Population: Patients with advanced CC, including patients after surgery or who are no longer candidates for surgical or liver-directed therapies.
- ❖ Intervention: First-line or greater-line (due to progression or toxicity) systemic therapy or stenting.
- ❖ Comparison: Interventions listed or placebo control.
- ❖ Outcomes: Primary outcomes were those explicitly reported as such in the Methods section of the published study or the registered trial in www.ClinicalTrials.gov. If none were explicitly reported, we considered the outcome used to calculate the sample size; if outcomes were not stated in the sample size calculation, the outcome defined in the primary study objective was retained. If the primary outcome was still not clearly identified, the article was excluded. Secondary outcomes were excluded from this study.

Data collection process

Using a piloted electronic data extraction form, a pair of reviewers (DA, NH) extracted data independently and in duplicate. All extracted data were verified by another reviewer (DB). For each eligible RCT, data relating to the primary outcome (i.e., number of events and nonevents for each arm, P value, and corresponding statistical tests) were extracted.

We also extracted the following data from each study: journal name, year of publication.

Calculation of SIFI

The SIFI was calculated from Kaplan-Meier (KM) curves, reconstructed using Digitizelt software (http://www.digitizeit.de/), by the iterative redesignation of the best survivors from the experimental to the control arm until statistical significance (defined as $P<0.05$) was lost. For trials with non-significant outcomes, negative SIFI was calculated similarly, but the direction was opposite—redesignation of the best survivors from the control to the experimental arm, until statistical significance was gained. Thus, the SIFI reflects statistical robustness, by denoting the minimal number of redesignations required to lose or gain

significance in survival analysis. P values were calculated with the 2-sided log-rank test. The follow-up time distribution was calculated using the 'prodlim' package in R.

Difference in RMST

To assess whether the experimental treatment has an effect on survival time without assuming proportional hazards, we calculated the absolute survival time of the patients in the participating studies. The RMST is a statistical calculation of the gain, or loss, in survival drawn from the area under the survival curve until a certain time point. When the difference in the RMST estimate is positive or negative, it is interpreted as a gain or loss of survival time, respectively.

Assessment of spin

In addition to the SIFI and the RMST, we quantified the level of distorted interpretation of study outcomes or "spin". The assessment of spin, was quantified as the number of sections with altered conclusions despite nonsignificant results, giving one point to each section with a spin on the study results, including the abstract, main text and conclusions.

When the primary outcome was not statistically significant, the level of spin was classified as high, moderate, or low/absent according to the classification described by Boutron *et al.* (3). When the primary outcome was statistically significant, the article conclusion provided by the authors was classified as providing recommendations for clinical practice.

We classified the level of spin based on specific criteria, as follows. High spin was noted when the study did not report possible issues in the study framing, when no clear recommendations for further investigations were suggested or when nonsignificant results for the primary outcomes were not highlighted as the main study outcome. A high level of spin was also noted when the author conclusions recommended using the suggested treatment in clinical practice despite nonsignificant results in the primary outcome. Moderate spin was defined as some uncertainty in the study framing or a recommendation for further trials with no acknowledgment of the statistically nonsignificant results for the primary outcomes. Low spin was noted as uncertainty in the study framing, recommendations for further trials, or acknowledgment of the statistically nonsignificant results for the primary outcomes.

Statistical analysis

The results are reported as the median and interquartile range (IQR) for continuous variables and as the number of occurrences with proportions represented as percentages for categorical variables. Hazard ratios were calculated using Cox proportional hazards regression models, and the statistical analysis was performed using the unadjusted log-rank test. All tests were two-sided. All statistical analyses were performed in R version 3.5.0.

Results

A total of 6167 articles were found in the initial search. After exclusion of nonrandomized controlled studies and review of the abstracts and full texts, fourteen randomized controlled trials remained. Three studies (4,5,9) were excluded due to the impossibility of calculating the SIFI due to a lack of appropriate survival curve figures, leaving 11 randomized trials (10-20) included in the final analysis. A PRISMA flowchart of the included studies is shown in Figure S1.

Nine studies investigated systemic chemotherapy, one study evaluated photodynamic therapy, and another investigated the use of an endoscopically inserted stent in the biliary tract. Gemcitabine and oxaliplatin (GEMOX), which is considered the standard therapeutic regimen for CC, was evaluated in four studies (11,13,14,16), while S-1-based chemotherapy was evaluated in three studies (15,17,19). Monoclonal antibodies as an addition to standard chemotherapy regimens were evaluated in two clinical trials (14,16). As for the study population, eight trials (12-17,19,20) included patients with metastatic/ locally advanced cancer, while 3 studies examined adjuvant therapy in patients following surgical resection (10,11,18). The median sample size for the eligible trials was 150 (IQR 75.5–246.5). The included studies are detailed in *Table 1*. The selected outcome was overall survival in 9 (81.8%) trials, and progression-free survival in 2 (18.2%) trials.

SIFI

The median SIFI across the included studies was –2 (IQR –6.25, –0.25). Three studies (17,18,20) with significant results had a positive fragility: 6 (ABC-02), 2 (BILCAP), and 2 (PDT trial), representing only 0.4–4.6% of the total sample size in these RCTs. A summary of the SIFI values of the studies can be seen in *Table 2* and an illustration of SIFI calculation is given in *Figure 1* (Figures on all studies

Table 1 Included studies

1st Author (ref)	Study ID and short name	Publication year, journal	Setting	No. of patients	Treatment vs. Control
Lee (12)	NCT01149122	2012, Lancet Oncology	Metastatic Disease, Both	268	GEMOX + Erlotinib vs. GEMOX
Leone (14)	NCT01389414 (Vecti-BIL)	2012, Cancer	Metastatic Disease, Both	89	GEMOX + Panitumumab vs. GEMOX
Li (15)	ChiCTR-TRC-14004733	2016, Oncotarget	Locally Advanced Disease, PHCC	75	GEM-S-1 vs. S-1
Malka (16)	NCT00552149 (BINGO)	2014, Lancet Oncology	Locally Advanced Disease, Both	150	GEMOX + Cetuximab vs. GEMOX
Valle (20)	NCT00262769 (ABC-02)	2010, New England Journal of Medicine	Metastatic/ Locally Advanced Disease, Both	410	Cisplatin + Gemcitabine vs. Gemcitabine
Primrose (18)	EudraCT2005-003318-13 (BILCAP)	2019, Lancet Oncology	Adjuvant Treatment, Both	447	Capecitabine vs. Best supportive care
Ebata (10)	UMIN000000820 (BCAT)	2018, British Journal of Surgery	Adjuvant Treatment, PH and distal CC	225	Gemcitabine vs. Best supportive care
Edeline (11)	EudraCT2008-004560-39 (PRODIGE 12-ACCORD)	2019, Journal of Clinical Oncology	Adjuvant Treatment, Both	196	GEMOX vs. Best supportive care
Sasaki (19)	UMIN000001703	2013, Cancer Chemotherapy and Pharmacology	Locally Advanced Disease, Both	62	GEM-S-1 vs. GEM
Park (17)	NCT00869635	2014, European Journal of Cancer	Locally Advanced Disease, PHCC	43	PDT + S-1 vs. PDT
Lee (13)	NCT02166970	2017, Gastrointestinal Endoscopy	Locally Advanced Disease, PHCC	133	Bilateral vs. Unilateral Stent

GEMOX, gemcitabine and oxaliplatin; PDT, photodynamic therapy; PH, peri-hilar; CC, cholangiocarcinoma. Both CC intra- and extrahepatic CC.

available on demand).

In addition, an example of SIFI calculation of the BILCAP trial can be seen in *Figure 2*.

RMST

Overall, the median RMST difference (RMST-D) for PFS (2 studies) and OS (9 studies) was 0.56 months (IQR: 0.10–0.95). The median RMST-D for overall survival was 0.6 months (IQR: 0.08–1.3). In two studies (11,19), the RMST-D was negative, indicating a loss in survival in the interventional group. The longest gain in survival was seen in the studies showing a significant improvement in overall survival using the log-rank test. One exception was seen in the trial that examined the effect of bilateral stenting in patients with hilar CC, where the RMST-D was significantly longer than that in patients treated with a

single stent, although the original overall survival analysis demonstrated nonsignificant results (P=0.056). In addition, the analysis of overall survival in the per-protocol group in the BILCAP trial (18), which was originally found to be statistically significant in favor of capecitabine (over observation), showed no significant advantage when the RMST-D was calculated. The RMST-D results of all included studies are shown in *Table 2*.

Spin effect

Assessment of the included studies with nonsignificant results demonstrated variable levels of spin in the abstracts and conclusions of all of the manuscripts. However, only in the BILCAP study (18) was spin applied to the study outcomes in all of the examined sections of the manuscript (abstract, main text, conclusion). Overall, the main text

Table 2 Main study outcomes

1st author (ref)	Primary endpoint	HR (95% CI); P value	SIFI	RMST (95% CI); P value
Lee (12)	PFS	0.80 (0.61, 1.03); 0.087	−1	0.83 (−0.13, 1.80); 0.091
Leone (14)	PFS	0.78 (0.51, 1.21); 0.27	−2	0.53 (−0.97, 2.02); 0.489
Li (15)	OS	0.68 (0.50, 0.90); 0.008	−2	0.72 (−0.51, 1.95); 0.252
Malka (16)	OS	NA	−7	0.60 (−0.53, 1.73); 0.296
Valle (20)	OS	0.64 (0.52, 0.80); <0.001	6	1.30 (0.55, 2.06); <0.001
Primrose (18)	OS (ITT)	0.81 (0.63, 1.04); 0.097	−3	0.08 (−0.21, 0.37); 0.587
Primrose (18)	OS (PPT)	0.75 (0.58, 0.97); 0.028	2	0.14 (-0.15, 0.43); 0.349
Ebata (10)	OS	1.01 (0.70, 1.45); 0.964	−15	0.11 (−0.17, 0.40); 0.432
Edeline (11)	OS	1.08 (0.70, 1.66); 0.74	−15	−0.40 (−0.85, 0.04); 0.077
Sasaki (19)	OS	NA	−6	−0.07 (−1.86, 1.72); 0.942
Park (17)	OS	0.36 (0.17, 0.75); 0.005	2	2.55 (1.07, 4.02); <0.001
Lee (13)	OS	0.679 (0.475, 1.009); 0.056	−1	1.71 (0.48, 2.94); 0.006

PFS, progression-free survival; OS, overall survival; ITT, intention to treat; PPT, per protocol treatment; NA, not available; SIFI, survival-inferred fragility index; RMST, Restricted mean survival time (in months).
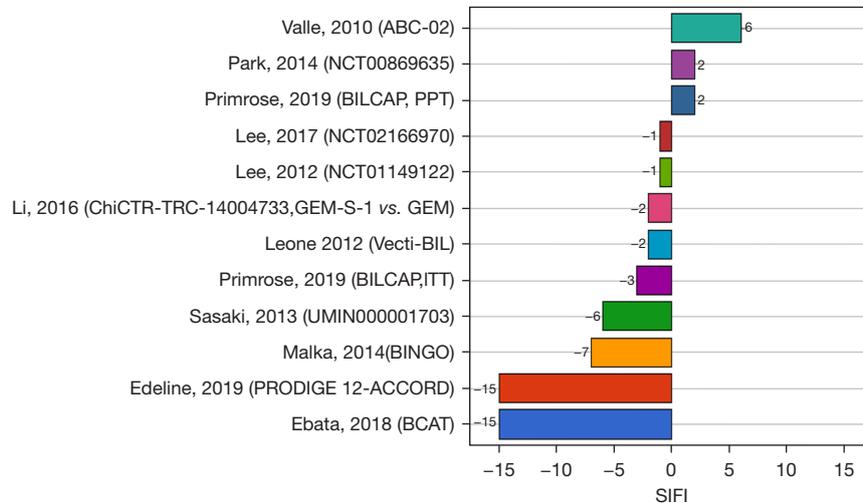
Figure 1 Survival-inferred fragility index (SIFI) of included studies.

was the most common location of spin within the paper (10-14,16,18,19) (8 studies), followed by the abstract (12,18,19) (3 studies) and the conclusion (16,18) (2 studies). In addition, a high level of spin was seen in one study (18), whereas a low level of spin was seen in five studies (10,11,13,14,19), reported mainly as a need for further investigations of the proposed treatment. The quantification of spin for the included studies is described in *Table 3*.

Discussion

Using additional statistical methods, the main messages of the present study are (I) across most studies, the SIFI analysis showed that the results are fragile and rely on a small number of patients to reverse statistical significance (or non-significance); (II) analysis of RMST-D showed that the median survival gained, if any, in the majority of studies was in the region of two weeks only; (III) the level of spin was

A

BILCAP (per-protocol analysis)
primrose 2019, lancet oncology

HR=0.75 (0.58,0.97); P=0.029

B

C

Baaelina
HR=0.75 (0.58, 0.97)
P=0.029

Iteration #1
HR=0.76 (0.59, 0.99)
P=0.039

Iteration #2
HR=0.77 (0.60, 1.00)
P=0.052
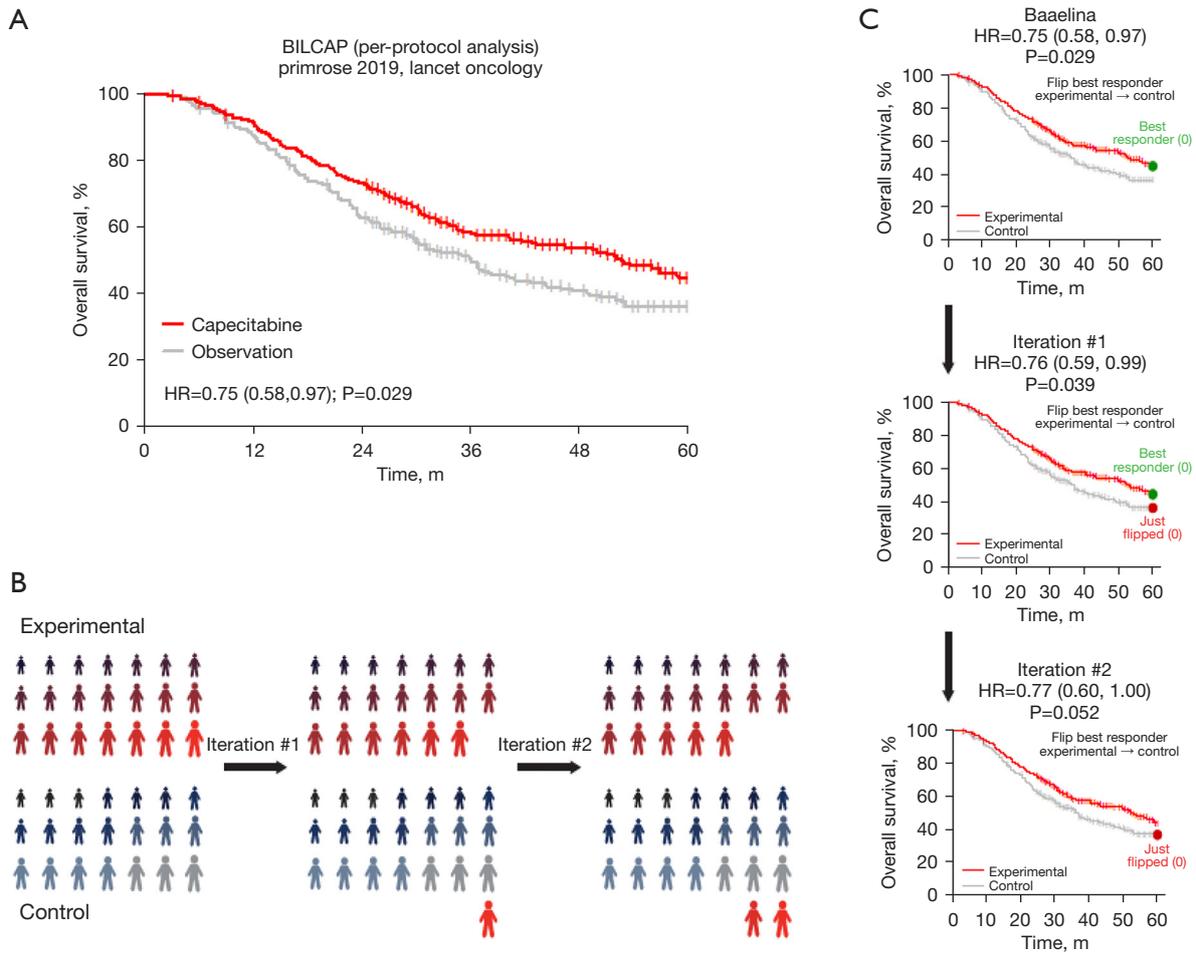
Figure 2 Example of survival-inferred fragility index (SIFI) calculation using BILCAP study.

Table 3 Evaluation of spin in RCTs of treatment of cholangiocarcinoma

1st Author (ref)	Abstract	Main text	Conclusion	Total
Lee (12)	High	Moderate	None	2
Leone (14)	None	Low	None	1
Malka (16)	None	Low	Low	2
Primrose (18)	High	Moderate	High	3
Ebata (10)	None	Low	None	1
Edeline (11)	None	Low	None	1
Sasaki (19)	Low	Low	None	1
Lee (13)	None	Low	None	1

Spin measured as per Boutron *et al.* (5).

high for inconclusive studies.

Several RCTs have tried to assess the role of systemic chemotherapy for CC, with contradictory results (10,11,18,20). Only a handful of randomized controlled studies showed any survival advantage due to adjuvant treatment in patients suffering from CC or in patients presenting with metastatic disease. Nonetheless, treatment guidelines still rely on these controversial results, and treatment regimens do not reflect the lack of supporting evidence (21).

The fragility index, first described by Walsh *et al.* (6), is a binary calculation of the number of patients responsible for a statistically significant finding in a randomized controlled trial. Although useful for binary outcomes, it is not applicable in time-to-event endpoints such as OS, DFS, PFS, etc. (8). The SIFI is an extension of the FI that allows to examine the robustness of the results from survival analyses.

A low absolute SIFI (a value that is close to zero), as seen in the majority of randomized trials examining the treatment of CC, indicates that the statistical significance of the survival analysis depends on a small number of patients, undermining the clinical significance in these studies, which also correlates with the poor outcome seen in the studies where a survival benefit was not demonstrated. Despite these results, practice parameters still include many treatments without significant evidence for successful outcomes, perhaps due to the lack of alternatives (21).

In addition, we evaluated the level of spin, as suggested by Boutron *et al.* (3). Spin, in which the scientific reporting of the study outcomes is filtered through the lens of the author's interpretation, is often misleading and can bring physicians ill-informed decision making. As seen in previous studies (22,23), there is often a discrepancy between the statistical significance of the study outcomes and the study conclusion and subsequent clinical recommendation. Adding the level of spin seen in these studies to the contradictory results and the low SIFI shows that although the majority of studies do not support the current treatment regimens, physicians are still urged to continue these treatments, despite the lack of meaningful efficacy.

Although statistical significance of study outcomes is commonly accepted as the required evidence for the incorporation of new treatments, it is important to note that statistical significance without adequate internal validity of the study methodology can lead to misinterpretation of study outcomes. For example, a recent comprehensive meta-analysis of published oncology RCTs by Rubinstein *et al.* (24), demonstrated that the majority of studies presented findings in a definitive manor, even when the P values suggested a substantial risk of a type I error. Similar observations were observed in RCTs for non-small cell lung cancer (25), breast cancer (26) and other malignancies (27). The ongoing debate about the interpretation of study results led some researchers to recommend withdrawing the dichotomous reporting of P values and use effect sizes and confidence intervals as a more appropriate way to report a clinically significant outcome (28,29).

Our study has several limitations which includes a limited number of eligible RCTs for this analysis and the selection of both intra- and extrahepatic CC or resectable and unresectable CC for evaluation. Furthermore, the rare incidence of CC poses a challenge in obtaining an adequate level of statistical power, especially when the effect size is small. One solution for this barrier is to encourage collaborations between researchers to conduct wide-ranging multicenter and multinational trials that would result in larger sample sizes and statistical power.

Our results, demonstrating a significant bias in the way randomized studies are conducted and reported, along with the dismal outcomes seen in patients suffering from CC, should encourage physicians to acknowledge the limitations of existing studies, to share the latter during the patient-doctor decision making, and more importantly to develop new robust strategies for the treatment of CC.

Acknowledgments

Footnote

8

Horesh et al. Robustness of the evidence for the management of CC

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Banales JM, Marin JJG, Lamarca A, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. Nat Rev Gastroenterol Hepatol 2020;17:557-88.
2. Gavriilidis P, Askari A, Roberts KJ, et al. Appraisal of the current guidelines for management of cholangiocarcinoma-using the Appraisal of Guidelines Research and Evaluation II (AGREE II) Instrument. Hepatobiliary Surg Nutr 2020;9:126-35.
3. Boutron I, Dutton S, Ravaud P, et al. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;303:2058-64.
4. Gruenberger B, Schueller J, Heubrandtner U, et al. Cetuximab, gemcitabine, and oxaliplatin in patients with unresectable advanced or metastatic biliary tract cancer: a phase 2 study. Lancet Oncol 2010;11:1142-8.
5. Boutron I, Altman DG, Hopewell S, et al. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. J Clin Oncol 2014;32:4120-6.
6. 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;67:622-8.
7. Bomze D, Asher N, Hasan Ali O, et al. Survival-Inferred Fragility Index of Phase 3 Clinical Trials Evaluating Immune Checkpoint Inhibitors. JAMA Netw Open 2020;3:e2017675.
8. Bomze D, Meirson T. A critique of the fragility index. Lancet Oncol 2019;20:e551.
9. Okusaka T, Nakachi K, Fukutomi A, et al. Gemcitabine alone or in combination with cisplatin in patients with biliary tract cancer: a comparative multicentre study in Japan. Br J Cancer 2010;103:469-74.
10. Ebata T, Hirano S, Konishi M, et al. Randomized clinical trial of adjuvant gemcitabine chemotherapy versus observation in resected bile duct cancer. Br J Surg 2018;105:192-202.
11. Edeline J, Benabdelghani M, Bertaut A, et al. Gemcitabine and Oxaliplatin Chemotherapy or Surveillance in Resected Biliary Tract Cancer (PRODIGE 12-ACCORD 18-UNICANCER GI): A Randomized Phase III Study. J Clin Oncol 2019;37:658-67.
12. Lee J, Park SH, Chang HM, et al. Gemcitabine and oxaliplatin with or without erlotinib in advanced biliary-tract cancer: a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2012;13:181-8.
13. Lee TH, Kim TH, Moon JH, et al. Bilateral versus unilateral placement of metal stents for inoperable high-grade malignant hilar biliary strictures: a multicenter, prospective, randomized study (with video). Gastrointest Endosc 2017;86:817-27.
14. Leone F, Marino D, Cereda S, et al. Panitumumab in combination with gemcitabine and oxaliplatin does not prolong survival in wild-type KRAS advanced biliary tract cancer: A randomized phase 2 trial (Vecti-BIL study). Cancer 2016;122:574-81.
15. Li H, Zhang ZY, Zhou ZQ, et al. Combined gemcitabine and S-1 chemotherapy for treating unresectable hilar cholangiocarcinoma: a randomized open-label clinical trial. Oncotarget 2016;7:26888-97.
16. Malka D, Cervera P, Foulon S, et al. Gemcitabine and oxaliplatin with or without cetuximab in advanced biliary-tract cancer (BINGO): a randomised, open-label, non-comparative phase 2 trial. Lancet Oncol 2014;15:819-28.
17. Park DH, Lee SS, Park SE, et al. Randomised phase II trial of photodynamic therapy plus oral fluoropyrimidine, S-1, versus photodynamic therapy alone for unresectable hilar cholangiocarcinoma. Eur J Cancer 2014;50:1259-68.
18. Primrose JN, Fox RP, Palmer DH, et al. Capecitabine compared with observation in resected biliary tract cancer (BILCAP): a randomised, controlled, multicentre, phase 3 study. Lancet Oncol 2019;20:663-73.
19. Sasaki T, Isayama H, Nakai Y, et al. A randomized phase II study of gemcitabine and S-1 combination therapy versus gemcitabine monotherapy for advanced biliary tract cancer. Cancer Chemother Pharmacol 2013;71:973-9.
20. Valle J, Wasan H, Palmer DH, et al. Cisplatin plus

gemcitabine versus gemcitabine for biliary tract cancer. N Engl J Med 2010;362:1273-81.

21. Network NCC. Hepatobiliary Cancers (Version 5.2020). 2020. Available online: https://www.nccn.org/professionals/physician_gls/pdf/hepatobiliary.pdf. Accessed August 4. 2020.

22. Chiu K, Grundy Q, Bero L. 'Spin' in published biomedical literature: A methodological systematic review. PLoS Biol 2017;15:e2002173.

23. Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. BMJ 2007;335:1202-5.

24. Rubinstein SM, Sigworth EA, Etemad S, et al. Indication of Measures of Uncertainty for Statistical Significance in Abstracts of Published Oncology Trials: A Systematic Review and Meta-analysis. JAMA Netw Open 2019;2:e1917530.

25. Sacher AG, Le LW, Leighl NB. Shifting patterns in the interpretation of phase III clinical trial outcomes in advanced non-small-cell lung cancer: the bar is dropping. J Clin Oncol 2014;32:1407-11.

26. Vera-Badillo FE, Shapiro R, Ocana A, et al. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. Ann Oncol 2013;24:1238-44.

27. Lazarus C, Haneef R, Ravaud P, et al. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. BMC Med Res Methodol 2015;15:85.

28. Bauchner H, Golub RM, Fontanarosa PB. Reporting and Interpretation of Randomized Clinical Trials. JAMA 2019;322:732-5.

29. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature 2019;567:305-7.

Identification

Records identified through database searching
(n=6,167)

Records excluded
(n=6,126)

Screening

Papers retrieved for evaluation of title and abstract
(n=41)

Papers excluded after review of
title and abstract (n=27)
- Phase I clinical trial (2)
- Study Protocol (3)
- Not RCT (8)
- Extrahepatic disease (13)
- Review Paper (1)

Eligibility

Full text evaluation for eligibility
(n=14)

Papers excluded for not
having a survival analysis
(n=3)

Inclusion

Number of papers included
(n=11)
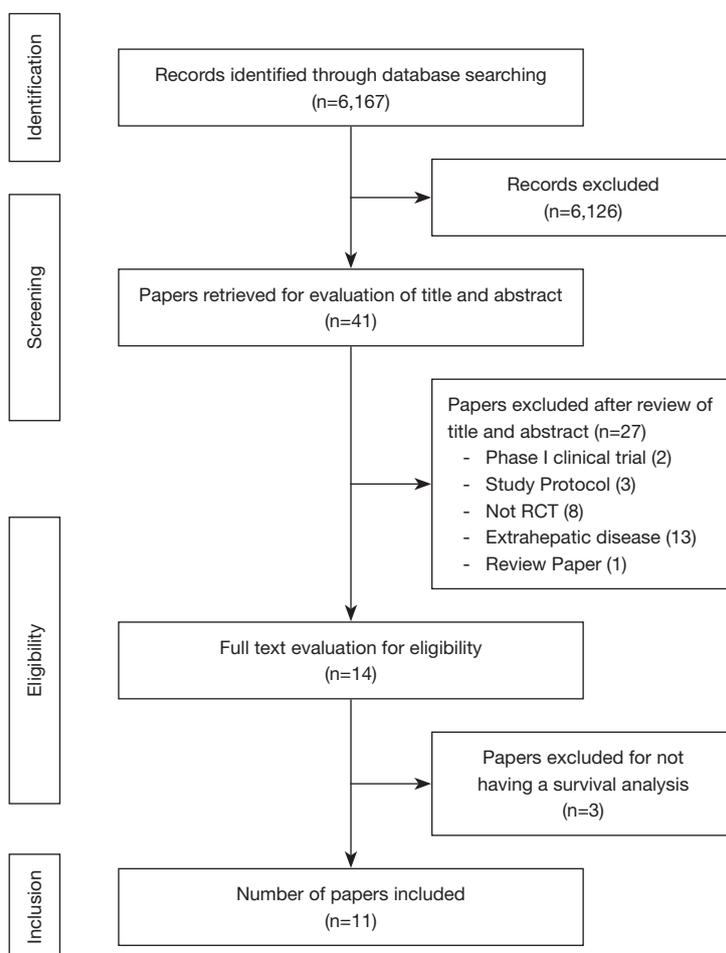
Figure S1 PRISMA flow diagram showing the search strategy along with the selection and screening processes for the eligible studies.