Research Article

Postoperative chemotherapy with capecitabine for biliary tract cancer? Quality and risk assessment of the BILCAP-study

Giulia Manzini[1,2*], Ursula Klotz[2], Ian N Hines[3], Doris Henne-Bruns[2], Michael Kremer[1], Marietta Kirchner[4]

[1]Department of General and Visceral Surgery, Hospital of Aarau, Tellstrasse 25, 5001 Aarau, Switzerland.

[2]Department of General and Visceral Surgery, University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany.

[3]Department of Nutrition Science, College of Allied Health Sciences, East Carolina University, Health Sciences Bldg. Room 4165F, Greenville, NC 27834, USA.

[4]Department of medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg.

[*]**Corresponding author:** Giulia Manzini, Department of General and Visceral Surgery, Hospital of Aarau, Tellstrasse 25, 5001 Aarau, Switzerland.

Received: 03 August 2021; **Accepted:** 10 August 2021; **Published:** 23 August 2021

Citation: Giulia Manzini, Ursula Klotz, Ian N Hines, Doris Henne-Bruns, Michael Kremer, Marietta Kirchner. Postoperative chemotherapy with capecitabine for biliary tract cancer? Quality and risk assessment of the BILCAP-study. Journal of Surgery and Research 4 (2021): 406-421.

Abstract

According to the ASCO guidelines, patients with resected biliary tract cancer should be offered adjuvant capecitabine chemotherapy based on the results of the BILCAP trial. The aim of this evaluation is to assess quality of the BILCAP study. The BILCAP study was analyzed according to the Delphi and the CONSORT checklists. Risk of bias was assessed using the Cochrane Risk of Bias Tool. On average, one patient was included every year by each center. The analysis was not adjusted for center. Treatment allocation by minimization was adopted but mode of application is poorly reported and the choice of variables not justified. No blinding was present. For the observed HR=0.81 with 234 events statistical power is only around 37%. Four of 9 items of the Delphi list and 6 of 35 items of the CONSORT checklist were not properly addressed. According to the Cochrane Risk of Bias Tool, the overall risk-of-bias judgement for the outcome overall survival of the BILCAP study was “some concerns”. Finally, the funding source had an advisory role in the study design. Based on the results of this study there is insufficient evidence for the administration of adjuvant chemotherapy with capecitabine in patients with biliary tract cancer.

Simple summary: This is a critical appraisal of the BILCAP study, which suggests a 6-months adjuvant capecitabine based chemotherapy after curative resection of biliary tract cancer. This study alone is the basis for the indication for adjuvant chemotherapy in the ASCO guidelines. Several pitfalls were found in the study, indicating that there is insufficient evidence for the administration of adjuvant chemotherapy with capecitabine in patients with biliary tract cancer.

Abbreviations:

ASCO: American Society of Clinical Oncology
HR: hazard ratio
ITT: Intention to treat
PP: per protocol
OS: overall survival
IQR: interquartile range
RFS: recurrence free survival
CTx: chemotherapy

Keywords: Liver surgery; Validity; Adjuvant chemotherapy; Cholangiocellular carcinoma; Gallbladder carcinoma

1. Introduction

Biliary tract cancers are classified as those associated with the intrahepatic bile ducts, perihilar and distal extrahepatic bile ducts, and the gallbladder [1]. A broad range exists in the five-year survival rates depending on location and stage of disease, from 2% to 15% and from 2% to 30% for intrahepatic and extrahepatic cholangiocarcinoma, respectively [2], and 2% to 70% for the gall bladder [3]. Resection is the primary therapeutic option for these patients though chemotherapeutic options are available including platin drugs and fluorouracil as well as more recently developed drugs including capecitabine and gemcitabine. Intriguingly, a very recent study highlighted the importance of resection plus non-surgical treatment as a mechanism to improve overall survival beyond chemotherapy alone [4]. Such combinatorial approaches are used in a variety of cancers though their efficacy in this specific setting have not been thoroughly evaluated. Capecitabine has received significant attention through its use as an adjuvant therapy in treatment of pancreatic adenocarcinoma. Acting as a pro-drug, capecitabine is metabolized to fluorouracil following oral administration leading to reduced DNA synthesis / repair and thus inhibition of tumor cell proliferation. Its use in other gastrointestinal tumors including those of the biliary tract have also been reported. Woo and colleagues [5] reported, in a retrospective study, a modest response of capecitabine when combined with cisplatin in advanced biliary tract cancer. Additionally, capecitabine in combination with oxiplatin has shown promise in phase II trial as secondary therapy following failure of gemcitabine and cisplatin [6]. Thus, approaches using capecitabine in combination with other traditional therapeutics improves multiple measures of disease progression within the biliary tract. The recently published BILCAP study (EudraCT number 2005-003318-13) aimed to determine the effectiveness of tumor resection in combination with adjuvant capecitabine chemotherapy in the setting of biliary tract cancer [7]. This two armed, randomized, controlled phase III study examined both recurrence free survival (RFS) as well as overall survival (OS) in 447 biliary tract cancer patients over an 8 years period undergoing curative resection with or without capecitabine treatment post-surgical intervention. The diagnosed patient population consisted of those with intrahepatic cholangiocarcinoma (19%), hilar cholangiocarcinoma (29%), gallbladder cancer (18%) and cholangiocarcinoma of the lower common bile duct (35%). Of the 447 total patients examined, 223 patients with biliary tract cancer resected with curative intent were randomly assigned to the capecitabine group and 224 to the observation group. The median follow-up for all patients was 60 months (IQR 37-60). The primary endpoint for the BILCAP study was OS. OS was not statistically different between groups in the intention-to-treat population (ITT) adjusting for minimization factors but ignoring center, which was the primary analysis of this trial. Median OS was 51.1 months (95%-CI 34.6-59.1) in the capecitabine group compared with 36.4 months (29.7-44.5) in the observation group (HR 0.81; 95%-CI [0.63, 1.04], p=0.097). In a sensitivity analysis in the ITT population, a difference in OS was found adjusted for minimization factors (ignoring center) and nodal status, disease grade and gender (HR 0.71; 95%-CI [0.55, 0.92]; p<0.01). A per-protocol analysis (PP) with 210 patients in the capecitabine group and 220 in the observation group also found a difference in median OS (53 vs. 36 months, HR 0.75; 95%-CI [0.58, 0.97]; p=0.028) in favor of capecitabine versus observation. At the time of the final analysis (March 6, 2017), 114 (51%), patients had died in the capecitabine group and 131 (58%) patients had died in the observation group. RFS also significantly favored the experimental group (HR 0.71; 95%-CI [0.54, 0.92], p=0.001). Overall, 280 (63%) of 447 patients had disease recurrence, 134 (60%) of 223 in the capecitabine group and 146 (65%) of 224 patients in the observation group. There was no significant difference in quality of life and the most common adverse event in the capecitabine group was palmar-plantar erythema [1]. The authors of the BILCAP study conclude that "although the trial was negative for the primary endpoint (OS by intention to treat), the data taken as a whole strongly suggest a benefit of adjuvant capecitabine" and that "capecitabine improves OS in the per-protocol population, with a clinically meaningful effect size of 14.7 months". Moreover, they state that "we believe this study is the first dedicated and sufficiently powered adjuvant study in biliary tract cancer and, as such, is uniquely placed to define the standard of care as capecitabine". Given the significant importance of these data and the potential to direct clinical practice in this area, the current study aims to thoroughly evaluate design, conduct, statistics and reporting of the BILCAP study providing an additional metric for overall method and study design quality as well an analysis of bias risk. To this end, the current evaluation subjected the BILCAP study to secondary analysis and highlighted important aspects of the study design which should be considered when determining overall potential effectiveness of this surgical + chemotherapeutic treatment strategy.

2. Methods

One approach in quality assessment is to focus on components such as randomization, blinding, allocation concealment, and sample size calculation in trial reports [8,9]. Another is to use a criteria list, for example, the list developed by Jadad et al. [10] and the Delphi list [11], to provide a quality score as an estimation of the overall methodological quality of the design and conduct of the trial [12,13]. Accordingly, design, conduct, and statistics of the BILCAP trial were analyzed focusing on single components as well as a comprehensive analysis using the Delphi list, which comprises a list of 9 questions to be answered with yes/no/do not know. The quality of the written report was assessed according to the CONSORT checklist. This checklist consists of 25 main items, several of them with sub-items, for a total of 37 individual points of review [14]. Two authors (GM and UK) analyzed the quality of the paper according to this checklist. In case of discordance, a third author (MK) provided a third review. The risk of bias was assessed using the RoB 2 Tool. This Cochrane Risk of Bias Tool was revised in August 2019 and considers bias arising at different stages of a trial, known as bias domains, which were chosen on the basis of both empirical evidence and theoretical considerations [15]. RoB 2 assessments relate to the risk of bias in a single estimate of intervention effect for a single outcome or endpoint, rather than for a whole trial as the risk of bias is outcome specific. The risk of bias for the primary endpoint of the BILCAP trial was assessed by two authors (GM and UK) according to the double check technique as described for the quality assessment according to the CONSORT checklist. In case of discordance, a third author (MKi) provided their assessment. RoB 2 is structured into five domains, bias arising from the randomization process, bias due to deviations from intended interventions, bias due to missing outcome data, bias in measurement of the outcome, bias in selection of the reported result. The risk-of-bias judgements for each domain are “low risk of bias”, “some concerns”, or “high risk of bias”. Judgments are based on, and summarize, the answers to signaling questions. The response options are “yes”, “probably yes”, “probably no”, “no”, and “no information”. According to the given answers, using an algorithm provided as supplement material here [15], is it possible to judge the risk for each domain. Thereafter, an overall risk-of-bias judgment can be provided for the study for a specific result (single outcome or endpoint, in this case overall survival) as follows: low risk of bias if the study is judged to be at low risk of bias for all domains for this result, some concerns if the study is judged to raise some concerns in at least one domain for this result, but not to be at high risk of bias for any domain and high risk of bias if the study is judged to be at high risk of bias in at least one domain for this result, or the study is judged to have some concerns for multiple domains in a way that substantially lowers confidence in the result [15].

3. Results

Several issues were identified in the BILCAP study regarding design, conduct, statistics and reporting and are summarized below.

3.1 Design

The BILCAP study is a randomized, controlled, multicenter phase 3 study which was performed across 44 specialist hepato-pancreato-biliary centers in the UK. Despite the inclusion of highly specialized centers, the number of included patients each year per center is extremely low. Specifically, a total of 447 patients were included by 44 centers over a period of 8 years, meaning that, on average, one patient was included in this study every year by each center. Patients were randomly assigned 1:1 to the capecitabine group or the observation group and allocation concealment was achieved using a computerized minimization algorithm that stratified patients by surgical center, site of disease, resection status, and performance status. However, minimization technique is poorly reported. Thus, it is not clear if and how a random element was introduced to make the allocation more unpredictable. The introduction of a random element into the procedure is, in fact, required by the ICH E9 guidelines, which state that "deterministic dynamic allocation procedures should be avoided and an appropriate element of randomization should be incorporated for each treatment allocation" [16]. Additionally, the program used to implement minimization is not reported. Blinding was not adopted as the control group did not receive any treatment. Inclusion of untreated controls limits the interpretation of the study. Specifically, the difference between the intervention and control group may be caused by a non-specific effect such as a placebo effect.

3.2 Conduct

Of the 447 patients, 280 (63%) had disease recurrence, 134 (60%) of 223 in the capecitabine group and 146 (65%) of 224 in the observation group). Follow-up treatment for patients who had disease recurrence was not recorded, leaving open the question regarding possible administration of CTx in 65% of the patients in the observation group. Different follow-up modalities are described for the control and intervention group as well which could be a source of bias. In particular, at the beginning of each treatment cycle, full blood count, biochemistry and liver function tests were done for the capecitabine group but not the control group. Baseline and periodic laboratory tests were done in both groups.

3.3 Statistics

The trial was negative for the pre-specified primary endpoint OS, analyzed in the ITT population, for which sample size was planned. Sample size calculation was based on the assumption that the 24-month OS would be 20% in the observation group and 32% in the capecitabine group, meaning 360 patients and 270 events needed to detect a hazard ratio (HR) of 0.71 with a 2-sided alpha level of 5% and a power of 80%. After a meeting of the independent data monitoring committee during the recruitment period, it became clear that the observed number of events was less than originally estimated. Therefore, it was recommended to do the final analysis once 234 events had accrued. Extensive power evaluations after adjusting the number of needed events, due to lower event rates than expected, were not done. It was argued that this number of events has a power of 80% to detect an HR of 0.69 so the study was deemed sufficiently powered. However, the observed effect in the primary analysis was only 0.81 which results in a power near 36% with 234 events and of around 41% with 270 events according to the Schoenfeld formula [17]. Additionally, with an HR=0.81, a median survival benefit of 15 months (36.4 vs. 51.1) was found in the capecitabine group, which would be of signfiicant clinical relevance. In oncological studies, an HR≤0.85 is considered clinically relevant, even if, for example, this reflects an improvement of OS of less than two months, as in the case of the RAISE trial. This randomized, double-blind, multicenter, phase 3 study compared ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that progressed during or after first-line therapy. Median OS was 13.3 months for patients in the ramucirumab group versus 11.7 months for placebo group with HR 0.844 (HR 0.844; 95%-CI [0.730, 0.976]; p=0.0219). This HR supported authorization of this second-line therapy [18]. The primary analysis and all subsequent sensitivity analyses were not adjusted by surgical center which was one of the minimization factors. The authors of the study stated that this is because of the large number of participating centers leading to flat statistical modelling regions. However, this result is not explained further and the possibility of running a frailty model with center included as a random effect is not mentioned which would be a solution to account for center in the situation of small number of patients per center. Moreover, treatment-by-center interactions were not addressed in exploratory analysis as suggested by the ICH E9 [16]. Adjusting for center in the analysis is generally recommended to obtain valid results [16]. It has been shown previously that, in unadjusted methods, the standard errors for treatment effect are biased upwards [19,20]. The authors mention that a prespecified sensitivity analysis was conducted in the ITT population where additional to the minimization factors the treatment effect was adjusted for further prognostic factors. However, in the statistical analysis plan published in the supplement, this analysis was not prespecified. The same is true for the study protocol where it is not clearly defined as a sensitivity analysis.

3.4 Quality assessment according to the Delphi list

3 out of 9 items of the Delphi list were non properly addressed (33.3%), these being the items regarding blinding (Table 1).

n.	Topic	Yes/no/don`t know
1	A) Was a method for randomization performed?	yes
	B) Was the treatment allocation concealed?	yes
2	Where the groups similar at baseline regarding the most important prognostic indicators?	yes
3	Where the elegibility criteria specified?	yes
4	Was the outcome assessor blinded?	no
5	Was the care provider blinded?	no
6	Was the patient blinded?	no
7	Were point estimates and measures of validity presented for the primary outcome measures?	yes
8	Did the analysis include an intention to treat analysis?	yes

Table 1: Assessment of the quality of the study according to the Delphi list [11]

3.5 Quality of the written report according to the CONSORT checklist

Authors explicitly state that they followed the CONSORT reporting guidelines, but they do not reference it. Six of 35 items of the CONSORT checklist were not properly addressed (17.1%) regarding randomization, sequence generation, allocation concealment, and blinding (Table 2). Again, the method of random sequence generation was poorly described making it difficult to determine if allocation concealment was maintained. The study was not blinded, although the outcome OS is less likely to be biased by this absence.

Section/Topic	*Item Number*	*BILCAP-Study*
Titel and Abstract	1a	yes
	1b	yes
Introduction		
Background and Objektives	2a	yes
	2b	yes
Methods		
Trial Design	3a	yes
	3b	yes
Participants	4a	yes
	4b	yes
Interventions	5	yes
Outcomes	6a	yes
	6b	yes
Sample Size	7a	yes
	7b	NA
Randomisation	8a	no
	8b	no
	9	no
	10	no
Blinding	11a	NA
	11b	yes
Statistical Methods	12a	yes
	12b	yes
Results		
Participant Flow	13a	yes
	13b	yes
Recruitment	14a	yes
	14b	yes
Baseline Data	15	yes
Numbers Analysed	16	yes
Outcomes and Estimaton	17a	yes
	17b	yes
Ancillary Analysis	18	yes
Harms	19	yes
Discussion		
Limitations	20	yes
Generalisability	21	no
Interpretation	22	no
Other Information		
Registration	23	yes
Protocol	24	yes
Funding	25	yes

NA*: not applicable

Table 2: Assessment of the quality of written report according to the CONSORT checklist [14].

3.6 Assessment of the Risk of bias of the study according to the RoB 2 tool

Overall survival was the endpoint analyzed with the RoB2 risk of bias tool. The risk assessment is presented in Table 3. Risk of bias was judged to be "low risk" for three of the five domains (bias arising from the randomization process, bias due to missing outcome data and bias in measurement of the outcome) and “some concerns” for the domains “bias due to deviations from intended interventions” and “bias in selection of the reported result”. Consequently, the overall risk of bias for the outcome OS was “some concerns” as the study is judged to raise some concerns in at least one domain. Regarding the first domain (bias arising from the randomization process), the general risk assessment was “low risk”. Even if the authors of the study describe to have incorporated a “random element” by using minimization, it is not possible to determine if the allocation sequence was specifically random. Regarding the second domain (bias due to deviations from intended interventions), even if the study is not blinded, the probability that this has affected the results of the study regarding the endpoint OS is minimal. For this reason, the risk of bias was classified as “some concerns” instead of “high” even without blinding. Regarding the domain "bias in selection of the reported result", as also reported from the authors of the BILCAP-study, there was no fully defined statistical analysis plan when the study was initiated, resulting in the risk judgment "some concern".

Bias domain and signalling question	**Response**
Bias arising from the randomisation process	
1.1 Was the allocation sequence random?	Probably Yes
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Probably yes
1.3 Did baseline differences between intervention groups suggest a problem with the randomisation process?	Probably no
Risk-of-bias judgement	Low risk
Bias due to deviations from intended interventions	
2.1 Were participants aware of their assigned intervention during the trial?	Yes
2.2 Were carers and people delivering the interventions aware of participants’ assigned intervention during the trial?	Yes
2.3 If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?	Yes
2.4 If Y/PY/NI to 2.3: Were these deviations likely to have affected the outcome?	Probably not
2.5 If Y/PY to 2.4: Were these deviations from intended intervention balanced between groups?	Not applicable
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Yes
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?	Not applicable
Risk-of-bias judgement	Some concerns
Bias due to missing outcome data	
3.1 Were data for this outcome available for all, or nearly all, participants randomised?	Yes
3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?	Not applicable
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	Not applicable
3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	Not applicable
Risk-of-bias judgement	Low risk
Bias in measurement of the outcome	
4.1 Was the method of measuring the outcome inappropriate?	No

4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	Probably no
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	yes
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	Probably not
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	not applicable
Risk-of-bias judgement	Low risk
Bias in selection of the reported result	
5.1 Were the data that produced this result analysed in accordance with a prespecified analysis plan that was finalised before unblinded outcome data were available for analysis?	No information
Is the numerical result being assessed likely to have been selected, on the basis of the results, from:	
5.2 ... multiple eligible outcome measurements (eg, scales, definitions, time points) within the outcome domain?	Probably yes
5.3 ... multiple eligible analyses of the data?	Probably yes
Risk-of-bias judgement	Some concerns

Table 3: Assessment of the Risk of bias of the study according to the RoB 2 tool [15]

3.7 Other shortcomings

A wide range of biological different tumor entities were included in the study (intrahepatic, hilar and low common bile duct cholangiocarcinoma as well as gallbladder carcinoma). It remains unclear if people treated with capecitabine have an increase in quality of life. Statistically significant differences were observed in the social functioning scale of the QLQ-C30 [21] in favor of the observation group (p= 0.006) as well as increased taste symptoms in the capecitabine group (p=0.042). Generally, even if not specifically clinically relevant, the QLQ-C30 functioning scale shows better results for the control group vs. the treatment group. Of the 28 authors, 7 (25%) declared to have received funds from several pharma industries, so a conflict of interest cannot be excluded. Additionally, the funding source for the study (Cancer Research UK and Roche) had an advisory role in design. Additionally, the first author of the BILCAP study was also involved in the generation of the ASCO guideline [1] as one of the 13 co-authors who developed this guideline. This guideline recommends that patients with resected biliary tract cancer should be offered adjuvant capecitabine CTx for a duration of 6 months. This recommendation is defined from the authors of the guideline as evidence based with an intermediate evidence quality and a moderate strength of recommendation [1]. The risk of bias of the BILCAP-trial was assessed using the Cochrane Risk of Bias Tool [15]. The BILCAP-trial was found to be at risk for bias due to lack of blinding of study participants and personnel. The survival outcomes of the BILCAP-study were rated in the ASCO guidelines as intermediate quality as a result of the significant magnitude of the overall survival effect in the per-protocol and prespecified adjusted ITT analyses. The main intention-to-treat analysis was not statistically significant. Despite the acknowledged presence of risk of bias in this study, the BILCAP-trial was the sole reference supporting treatment with adjuvant capecitabine CTx in the ASCO guidelines.

4. Discussion

This paper is a critical analysis of the BILCAP study [7] which found, in the pre-specified sensitivity and per-protocol analyses, an improvement of overall survival with capecitabine in patients with resected biliary tract cancer when used as adjuvant chemotherapy following surgery in comparison to observation. According to the results of the BILCAP study, adjuvant capecitabine-based chemotherapy is offered to patients after surgical resection. It is generally important to critically assess quality and bias of studies because both jeopardize validity. Poor quality studies are often included in meta-analysis as well as taken as basis for national and international guidelines [22-25], with the consequence that patients are potentially offered therapies which may not be effective in terms of improvement of overall survival and may even dramatically reduce the quality of life. Only 55% of the patient in the intervention group could complete the eight cycles of CTx. 32% discontinued the capecitabine-based therapy because of toxicity. Quality of RCTs has been defined as “the likelihood of the trial design to generate unbiased results” [10]. As this definition covers only the dimension of internal validity, Verhagen et al. proposed in 2001 the following definition of quality; “the likelihood of the trial design to generate unbiased results, that are sufficiently precise and allow application in clinical practice”, which comprises internal validity, external validity and statistical analysis [12]. Bias is a systematic deviation from the effect of intervention that would be observed in a large randomized trial without any flaws [15]. Quality can include study characteristics such as performing a sample size calculation that are not inherently related to bias in the study`s results [15]. In addition to quality, the risk of bias of the BILCAP study were also assessed. In particular, quality was assessed according to the Delphi list [11] and CONSORT checklist [14], while the risk of bias was assessed according to the new version of the Cochrane Risk of bias tool [15]. The first limitation of this study is the lack of statistical significance of the treatment effect for the primary endpoint in the primary ITT analysis. As stated by the authors, "although the trial was negative for the prespecified primary endpoint (overall survival by intention to treat), the data take as whole strongly suggest a benefit of adjuvant capecitabine". The primary analysis, on which a statistical significance relies, should be done on the ITT population because, in this population, the known and unknown confounder are equally distributed according to the randomization procedure and potential bias due to exclusion of patients is avoided [16]. In the sensitivity and per protocol analysis, which were statistically significant, it cannot be judged if the balance is maintained or that enough power was present. Consequently, these analyses are more likely to be biased. However, the true HR is not known. For a HR=0.81, power is only 36% for 234 events and 41% for 270 events. To achieve a power of 80%, 700 events would be necessary. Considering that the authors state, “we believe this study is the first dedicated and sufficiently powered adjuvant study in biliary tract cancer, and, as such, is uniquely placed to define the standard of care as capecitabine”, our power calculations raise concerns about the statistical power of the BILCAP study. The question if a statistically significant result not in the primary analysis is reliable enough to justify the recommendation in the ASCO guidelines remains open. As the true effect is not known and the different analyses yield different results with respect to the observed treatment effect, it would be important to discuss these differences to get a better idea of the true treatment effect. The larger effect in the sensitivity analysis when adjusting for further covariates is not explained. Likely, problems of overfitting arise here due to the large number of covariates present in the model. A minimization method [26-29] is used which is a type of dynamic allocation aiming at the achievement of a balance with respect to a large number of pre-specified prognostic factors. Here, the new subject’s treatment assignment is determined by investigating the potential covariate imbalance that would result if the subject was assigned to the treatment or control group, respectively [30]. Concerns are raised over this design as it compromises adequate generation of an allocation sequence and concealment in this study. By using minimization, investigators can determine the group to which a prospective subject would be allocated and then decide whether this is positive or negative in terms of creating an imbalance in some key predictor of outcome not considered in the imbalance function. Adding randomization, which means that the treatment that minimizes the imbalance function is not necessarily allocated, does not fully solve this issue [31]. The European Medicines Agency´s (EMA) Committee [CPMP] states that “dynamic allocation is strongly discouraged” [32]. The primary analysis was adjusted for minimization factors but not for centers. Authors explain that this is because of the high number of participating centers (n=44) and are aware that this is a limitation of the study. The option of including centers as a random effect was not considered. Actually, this could have been a disadvantage as adjusting for stratification factors can lead to an increase in power because the standard errors are not biased upwards and consequently confidence intervals become narrower [20,33]. Thus, to yield correct inference, it is necessary to include all minimization factors in the analysis. Moreover, according to the ICH E9 guidelines [16] in the exploratory analysis, center x treatment interactions as subgroups should be considered. In the original study, nothing is written to address the factor “center”, only “….not adjusted for surgical center because of large number leading to flat statistical modelling regions”. This is difficult to understand from a statistical point of view as only specialized centers in hepato-pancreato-biliary surgery participate. It is remarkable that, on average, each center could recruit only one patient and this could indicate selection bias. A long recruitment time (8 years) makes the comparison between patients difficult, as the surgical technique as well as the instruments used evolve and likely affects the mortality rates independent of specific treatments examined in this study. Additionally, despite the extremely low number of included patients in each center, wide inclusion criteria were adopted for this study. Patients with intrahepatic cholangiocarcinoma, hilar cholangiocarcinoma, gallbladder carcinoma and lower common bile duct cholangiocarcinoma were included. These tumor entities are very heterogeneous making it difficult to recognize if capecitabine can have a beneficial survival effect in a specific tumor type. Future well powered RCT should focus on a particular tumor entity. The lack of a placebo-controlled and blinded study affects the validity of this study. Without placebo control, it is impossible to differentiate between specific pharmacological and placebo effects, even if overall survival is less likely to be affected from a placebo effect than other endpoints, like quality of life and pain [1,34]. Placebo effect is defined as the “response of a subject to a substance or any procedure known to be without specific therapeutic effect for the condition being treated” [35]. Several studies demonstrated that perceptual characteristics of drugs [36], the route of administration [37], laboratory tests [38], diagnosis [39], and doctor-patient relationship play an important role in the outcome of illness [40-43]. Information regarding treatment or no treatment alone is sufficient to cause a placebo effect [44]. Moreover, patients´ and doctors´ preferences could also have influenced the results in an open study [45]. Patients assigned to the control group feel disadvantaged because they expect to be treated. Furthermore, when there is no concealment of treatment allocation, the randomization procedure is compromised because of conscious or subconscious bias [46]. The use of different follow-up modalities could be a source of bias because any treatment and additional attention from the doctor (difference in care) could lead to an improvement in the patients´ outcome [47]. Moreover, Sox et al. [48] found that laboratory tests that have no diagnostic value were independent factors of recovery. Follow-up treatment for patients who had disease recurrence was not recorded, leaving open the question about the possible administration of CTx in 65% of the patients in the observation group. It is realistic to suppose that patients with local recurrence received palliative CTx. Consequently, overall survival is questionable as an appropriate primary endpoint. Recurrence free survival could have been chosen as the primary endpoint and overall survival as secondary. The QLQ-C30 functioning scale showed generally better results for the intervention group, even if not statistically significant, returning to the question if patients should be offered a "moderate", evidence-based therapy based on an "at risk of bias" study which is not statistically significant in the ITT analysis, on which sample size was calculated. When the quality of the study is assessed with the Delphi list, approximately 50% of the items are not treated properly, specifically those pertaining to blinding and allocation concealment. The BILCAP study was reported according to the CONSORT checklist, where 6 items were not properly reported. A flawed report (i.e., lacking the necessary information regarding the trial) does not necessarily mean that the underlying study was flawed [12]. It is also true that a good written report does not automatically implicate good design, conduct and statistics within the study. Regarding the risk of bias assessment with the RoB 2 Tool, the main problems centered around the randomization procedure and the lack of a pre-specified statistical analysis plan. Consequently, the judgement involving the risk of bias of the BILCAP study was "some concerns". Following the results of the BILCAP study and according to the ASCO guidelines, adjuvant capecitabine should be offered to patients after resection of biliary tract cancer. However, it is important to note that the ASCO guidelines rely on a unique study which contains several limitations. In particular, as described in the ASCO guidelines, the BILCAP study is "at risk of bias" and the strength of the recommendation for the adjuvant chemotherapy is low. No other trials other than BILCAP have demonstrated an advantage in survival of any adjuvant CTx in resected bile duct cancer. In 2018, the randomized phase III clinical trial of adjuvant gemcitabine CTx versus observation in resected bile duct cancer (BCAT study) was published. A total of 225 patients were included (117 in the intervention group and 108 in the observation group). The study found no significance differences in OS (median 62.3 versus 63.8 months, respectively; HR 1.01 with 95%CI (0.70;1.45), p=0.964) and relapse-free survival (median 36 versus 39.9 months; HR 0.93, 95%CI (0.66;1.32), p=0.693). Even in the subgroup analysis after stratification for lymph node status and margin status, no survival difference between the two groups was observed. Toxicity was higher in the intervention group. This study concluded that the survival probability in patients with resected bile duct cancer was not significantly different between the intervention and the observation group [49]. The PRODIGE 12 study is a randomized multicenter phase III trial aimed to assess whether gemcitabine and oxaliplatin CTx would increase relapse-free survival while maintaining health-related quality of life in patients who undergo resection. A total of 196 patients were included in the study. No significant difference in relapse free survival between the two arms (median, 30.4 months in the intervention group versus 18.5 months in the control group; HR 0.88; 95%CI (0.62;1.25), p=0.48) as well as in overall survival (median, 75.8 months in the intervention group versus 50.8 months in the control group; HR 1.08, 95%CI (0.70;1.66), p=0.74) was observed. Toxicity was higher in the intervention group (p<0.001) [50].

5. Conclusion

In conclusion, several pitfalls were identified in the BILCAP study, which is routinely cited at tumor conferences to justify the administration of capecitabine after curatively resected biliary tract cancer. This remains the only study reporting an advantage for any form of adjuvant CTx, even in the presence of high toxicity for the patients. Definitive assessments of this topic should be delayed until future trials are properly developed.

Author Contributions

Author´s contributions: GM, MKi, MK and DHB contributed substantial to conception and design of the study, GM, INH and MK contributed in analysis and interpretation of the data. GM and UK assessed the quality of the study as well as the risk assessment. By inconsistencies MKi was asked. GM and MKi drafted the article. MKi revised critically the statistical aspects of the study and perform power calculations. All authors gave the final approval of the version to be published.

Funding

This research received no external funding.

Institutional review Board Statement

Not applicable.

Informed Consent Statement

Not applicable. No subject are involved in this work, which is a quality and risk assessment of an already published study.

Data availability statement

Not applicable. No data are analyzed in this publication.

Acknowledgment

None

Conflicts of interest

The authors declare no conflict of interest.

References

1. Shroff RT, Kennedy EB, Bachini M, et al. Adjuvant therapy for resected biliary tract Cancer: ASCO clinical Practice Guideline. J Clin Oncol 37 (2019): 1015-1027.

2. American Cancer Society: Bile duct cancer stages (2021).

3. Hundal R, Shaffer EA. Gallbladder Cancer: Epidemiology and outcome. Clin Epidemiol 6 (2014):

99-109.

4. Yin L, Zhao S, Zhu H, et al. Primary tumor resection improves survival in patients with multifocal intrahepatic cholangiocarcinoma based on a population study. Scientific Reports 11(2021): 12166.

5. Woo SM, Lee WJ, Han SS, et al. Capecitabine plus cisplatin as first-line chemotherapy for advanced biliary tract cancer: a retrospective single-center study. Chemotherapy 58 (2012): 225-232.

6. Kim TK, Oh SY, Lee J, et al. Capecitabine plus Oxaliplatin as a Second-Line therapy for Advanced Biliary Tract Cancers: A Multicenter, Open-Label, Phase II Trial. J Cancer 10 (2019): 6185-6190.

7. Primrose JN, Fox RP, Palmer DH, et al. Capecitabine compared with observation in resected biliary tract cancer (BILCAP): a randomized, controlled, multicentre, phase 3 study. Lancet Oncol 20 (2019): 663-73.

8. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. Intern J Tech Asses Health Care 12 (1996): 195-208.

9. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273 (1995): 408-412.

10. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin trials 17 (1996): 1-12.

11. Verhagen AP, De-Vet HCW, De-Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conductiong systematic reviews developed by Delphi consensus. J clin Epidemiol 51 (1998): 1235-1241.

12. Verhagen AP, De Vet HCW, De Bie RA, et al. The art of quality assessment of RCTs included in systematic reviews. Journal of Clinical Epidemiology 54 (2001): 651-654.

13. De Bie RA. Methodology of systematic reviews: an introduction. Phys Ther Rev 1 (1996): 47-51.

14. Moher D, Schulz KF, Altman DG. CONSORT GROUP (Consolidated Standards of Reporting Trials). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med 134 (2001): 657-662.

15. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366 (2019): l4898.

16. ICH E9: Statistical Principles for Clinical Trials. London UK: International Conference on Harmonization CPMP/ICH/363/96 (1998)

17. Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distributions. Biometrika 68 (1981): 316-319.

18. Tabernero J, Yoshino T, Cohn AL, et al. Ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that progressed during or after first-line therapy with bevacizumab, oxaliplatin, an fluoropyrimidine (RAISE): a randomized, double-blind, multicenter, phase 3 study. Lancet Oncol; 16 (2015): 499-508.

19. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. BMJ: British Medical Journal (2012).

20. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. Statistics in Medicine 31 (2012): 328-

340.

21. QLQ-C30, available at qol.eorct.org (2020).

22. Manzini G, Henne-Bruns D, Kremer M. Validity of studies suggesting post-surgical chemotherapy for resectable gastric cancer: critical appraisal of randomized trials. BMJ Open Gastroentol 4 (2017).

23. Manzini G, Klotz U, Henne-Bruns D, et al. Validity of studies suggesting preoperative chemotherapy for resectable thoracic esophageal cancer: A critical appraisal of randomized trials. World J Gastrointest Oncol 12 (2020): 113-123.

24. Manzini G, Hapke F, Hines I, et al. Adjuvant chemotherapy in curatively resected rectal cancer: How valid are the data?. World J Gastrointest Oncol 12 (2020): 503-513.

25. Manzini G, Hines IN, Henne-Bruns D, et al. Resection or radiofrequency ablation for hepatocellular carcinoma? Assessment of validity of current studies, meta-analyses and their influence on guidelines. HPB (Oxford) S1365-182X(19)33228-9 (2020).

26. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31 (1975): 103-115.

27. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. Clinical Pharmacology and Therapeutics 15 (1974): 443-453.

28. Wei LJ. A class of designs for sequential clinical trials. Journal of the American Statistical Association 72 (1977): 382-38626.

29. Wei LJ. The adaptive biased coin design for sequential experiments. The Annals of Statistics 6 (1978): 92-100.

30. Xu Z, Proschan M, Lee S. Validity and power considerations on hypothesis testing under minimization. Statistics in Medicine 35 (2016): 2315-2728.

31. Berger VW. Minimization, by its nature, precludes allocation concealment, and invites selection Bias. Contemp Clin Trials 31 (2010): 406.

32. Committee for Proprietary Medicinal Products (CPMP). Committee for proprietary medicinal products (cpmp): points to consider on adjustment for baseline covariates. Statistics in Medicine 23 (2004): 701-709.

33. Kahan BC, Morris TP. Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects?

34. Chvetzoff G, Ian F. Tannock IF. Placebo effects in Oncology. JNCI 95 (2003): 19-29.

35. Benedetti F, Amanzio M. The neurobiology of placebo analgesia: from endogenous opiois to cholecystokinin. Progr Neurobiol 52 (1997): 109-125.

36. Buckalew LW, Coffield KE. An investigation of drug expectancy as a function of capsule color and size and preparation form. J Clin Psychopharmacol 2 (1982): 245-248.

37. Wall PD. Pain and the placebo response. Ciba Found Symp 174 (1993): 187-216.

38. Sox HC Jr, Margulies I, Sox CH. Psychologically mediated effects of diagnostic tests. Ann Intern Med 95 (1981): 680-685.

39. Thomas KB. General practice consultations: is there any point in being positive? Br Med J (Clin Res Ed); 294 (1987): 1200-1202.

40. Bass MJ, Buck C, Turner L (1986). The physician´s actions and the outcome of illness in family practice. J Fam Pract; 23 (1986): 43-47.

41. Gracely RH, Dubner R, Deeter WR et al. Clinician´s expectations influence placebo analgesia. Lancet 1 (1986):1-43.

42. Greenfield S, Kaplan S, Ware JE Jr.

Expanding patient involvement in care. Effects on patient outcomes. Ann Intern Med 102 (1985): 520-528.

43. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ 152 (1995): 1423-1433.

44. Waber RL. Commercial Features of Placebo and Therapeutic Efficacy. JAMA 299 (2008): 1016-1017.

45. Porzsolt F, Eisemann M, Habs M et al. Form follows function: pragmatic controlled trials (PCTs) have to answer different questions and require different designs than randomized controlled trials (RCTs). J Public Health 21 (2013): 307-313.

46. Altman DG, Schulz K. Concealing treatment allocation in randomized trial. BMJ 323 (2001): 446-447.

47. Koshi EB, Short CA. Placebo Theory and Its Implications for Research and Clinical Practice: A Reviwe of the Recent Literature. Pain Practice 7 (2007): 4-20.

48. Sox HC Jr, Margulies I, Sox CH. Psychologically mediated effects of diagnostic tests. Ann Intern Med 95 (1981): 680-685.

49. Ebata T, Hirano S, Konishi M, et al. Randomized clinical trial of adjuvant gemcitabine chemotherapy versus observation in resected bile duct cancer. BJS 105 (2018): 192-202.

50. Edeline J, Benabdelghani M, Bertraut A, et al. Gemcitabine and Oxaliplantin Chemotherapy in Resected Biliary Tract Cancer (PRODIGE 12-ACCORD 18-UNICANCER GI): A Randomized Phase III study. JCO 37 (2020): 658-667.