Research Article
Hepatic and Biliary Cancer

JOURNAL OF HEPATOLOGY

Association of candidate surrogate endpoints with overall survival in advanced biliary tract cancer

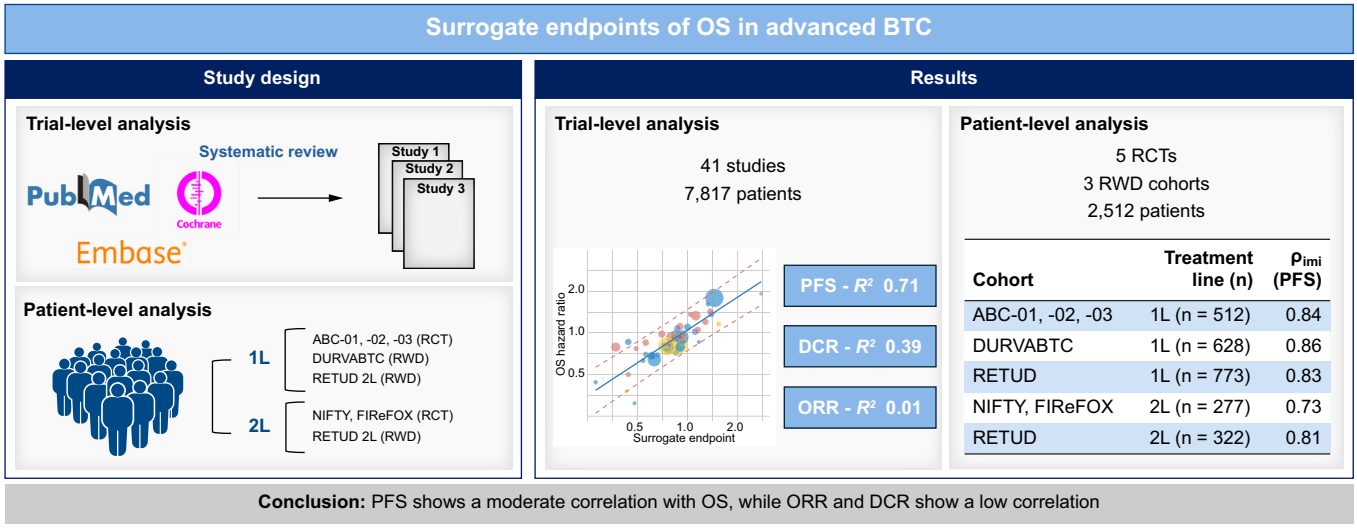
Authors

Florian Castet, Carles Fabregat-Franco, John Bridgewater, …, Changhoon Yoo, Juan W. Valle, Teresa Macarulla

Correspondence

tmacarulla@vhio.net (T. Macarulla).

Graphical abstract



Highlights

- PFS showed a moderate correlation with OS at the trial- and patient-level.
- A PFS hazard ratio of 0.61 in a hypothetical trial of 200 patients would likely lead to an OS benefit.
- Disease control rate and response rate showed a low correlation at the trial-level.
- Patients who responded to first- or second-line chemotherapy did not show significantly improved OS.

Impact and implications

The use of validated surrogate endpoints in biliary tract cancer trials may decrease costs and improve study feasibility, particularly with agents that only target small subsets of patients or in trials that incorporate a crossover design. A formal statistical validation of surrogacy requires patient-level and trial-level data. This is the first comprehensive analysis to incorporate novel agents (including immunotherapies and targeted agents), include patient-level data and rigorously and homogeneously extract appropriate measures of treatment effect for endpoint correlation. These results show a moderate correlation for progression-free survival both at the trial- and patient-level and a low correlation for disease control rate and response rate. This information will aid clinicians in appropriately interpreting contemporary clinical trials and guide clinical researchers and trial sponsors involved in clinical trial design. Furthermore, it has important implications for the regulatory approval process and may aid agencies in appropriately evaluating novel drugs.

https://doi.org/10.1016/j.jhep.2025.05.020

Research Article
Hepatic and Biliary Cancer

JOURNAL
OF HEPATOLOGY

Association of candidate surrogate endpoints with overall survival in advanced biliary tract cancer

Florian Castet[1], Carles Fabregat-Franco[2], John Bridgewater[3], Jin Won Kim[4], Margherita Rimini[5], Adelaida La Casta[6], Angela Lamarca[7,8], Minsu Kang[4], Francesca Salani[9], Alfredo Castillo[10], Andre Lopes[11], Jaewon Hyung[12], Lorenza Rimassa[13,14], Jorge Adeva[15], Daniel López-Valbuena[1], Míriam Basagaña-Farres[16], Simran Vaja[11], Ka Man Mak[11], Tian V. Tian[1], Andrés Muñoz[17], on behalf of the Spanish Cooperative Group for the Treatment of Digestive Tumors (TTD), Andrea Casadei-Gardini[5], on behalf of the DURVABTC Group, Changhoon Yoo[12], Juan W. Valle[18,19], Teresa Macarulla[1,*]

Journal of Hepatology **2025**. vol. 83 | 1102–1115

Check for updates

See Editorial, pages 1001–1003

Background & Aims: Surrogate endpoints are increasingly used in biliary tract cancer (BTC) trials. While this may expedite drug approval and decrease costs, surrogate endpoints may not always correlate with an overall survival (OS) advantage. We aimed to explore the association of progression-free survival (PFS), objective response rate (ORR) and disease control rate (DCR) with OS at the trial- and patient-level.

Methods: For the trial-level analysis, we performed a systematic review of Pubmed/MEDLINE, Embase, Cochrane, clinicaltrials.gov and conference proceedings for phase II-III trials in advanced BTC. We used a weighted linear regression to measure the correlation of OS with PFS, ORR and DCR. For the patient-level analysis, we analyzed patients included in five randomized trials and three real-world datasets. The protocol is registered with PROSPERO, CRD42023398279.

Results: For the trial-level analysis, we included 41 studies, involving 88 treatment arms and 7,817 patients. The coefficient of determination (R^2) of the model was 0.71 (95% CI 0.56-0.86) for PFS, 0.01 (0-0.08) for ORR and 0.39 (0.14-0.64) for DCR. Predefined subgroup analysis showed consistent results. For the patient-level analysis, we included a total of 2,506 patients, 783 in randomized trials (first-line 512, second-line 271) and 1,723 in routine clinical care (first-line chemotherapy 773, first-line chemotherapy-durvalumab 628, second-line chemotherapy 322). Across the distinct datasets, the correlation coefficient ranged from 0.73 to 0.86 for PFS. A responder analysis found no association between response and survival.

Conclusion: PFS shows a moderate correlation with OS both at the trial- and patient-level, while ORR and DCR show a low correlation. Whilst PFS is currently the best candidate surrogate marker for OS, our results highlight the need for novel endpoints in this field.

Introduction

Biliary tract cancers (BTCs) are a heterogeneous group of aggressive neoplasms arising in the biliary tree.[1] Around 60% of tumors are diagnosed at advanced stages and more than 70% of tumors treated with local curative treatments will eventually relapse,[2] resulting in a dismal median survival of about 1 year despite optimal systemic treatment.[3,4]

In this setting, overall survival (OS) is the most robust, reliable and clinically meaningful endpoint for the design of randomized-controlled trials (RCTs).[5] The relatively low follow-up necessary to reach sufficient events, coupled with the scarcity of effective treatment options beyond first-line therapy, make OS an ideal endpoint and less prone to biases arising from post-progression treatment imbalances and biological differences in molecular subgroups.[6,7] However, some circumstances may hinder the interpretation of OS, such as crossover designs or conditional accelerated approval programs, where the experimental drug is made available to clinicians during the execution of the validation trial, leading to uncontrolled post-progression crossover.

Surrogate endpoints are intended to substitute for final patient-relevant outcomes that directly measure how patients feel, function or survive in clinical trials.[8] The use of surrogates is cost-effective and may overcome some of the challenges associated with OS. The use of these endpoints in oncology trials has increased dramatically in recent years, best reflected by the fact that 78% of drug approvals by the US FDA

* Corresponding author. Address: Vall d'Hebron Institute of Oncology, C/ Natzaret, 115-117 08035 Barcelona, Spain; Tel.: 934894350, fax: 932746781.
E-mail address: tmacarulla@vhio.net (T. Macarulla).
https://doi.org/10.1016/j.jhep.2025.05.020

between 2005 and 2023 were based on surrogate endpoints.[9] However, only 32% of approved indications based on surrogate endpoints eventually demonstrated an improvement in OS,[10] highlighting the need for appropriate validation of these endpoints.

Previous studies have explored the association of progression-free survival (PFS) and objective response rate (ORR) with OS in BTC, although the results have been conflicting, the statistical methodology has been suboptimal, and only trial-level information has been included.[11–14] Despite the lack of robust data supporting the use of surrogate endpoints in BTC, 25% of randomized phase II-III trials used ORR as a primary endpoint and 44% used PFS.[15] In addition, the FDA has granted accelerated approval for pemigatinib and futibatinib based on ORR and duration of response for *FGFR2*-rearranged tumors and regular approval to ivosidenib based on PFS for *IDH1*-mutant tumors.[16]

To address these issues and explore the feasibility of using surrogate endpoints in advanced BTC, we performed a comprehensive analysis evaluating the association of PFS, ORR and disease control rate (DCR) at a trial-level through a meta-analysis of RCTs and at a patient-level through an analysis of five cohorts comprising both patients treated within the context of a RCT and in the real-world setting.

Patients and methods

Theoretical framework

One of the most used methodologies for evaluating potential surrogate endpoints is the two-stage meta-analytic framework, which requires individual patient-level data from all trials included in the systematic review to calculate the individual- and trial-level correlation.[17] In this framework, a validated endpoint will meet two conditions: demonstrate a correlation of treatment effects on both endpoints (Condition 1) and a strong correlation between the surrogate and definitive endpoint (Condition 2). One of the major limitations of this approach is that identified trials whose individual data cannot be retrieved are excluded from the trial-level analysis, which leads to a selection bias.[18,19] To address this potential limitation and because we did not have access to individual-level data from all identified trials in the systematic review, we applied an adaptation of this framework that intended to demonstrate both conditions.

For Condition 1 (trial-level), we performed a systematic review and correlation analyses of all trials based on aggregate-level data, as detailed below. For Condition 2 (patient-level), we analyzed two cohorts of patients included in RCTs evaluating first-line (ABC-01,[20] ABC-02[21] and ABC-03[22]) and second-line chemotherapy (NIFTY[23] and FIReFOX[24]). Given the complementary information provided by real-world data (RWD),[25] especially in the context of PFS, which is sensitive to the timing of assessments and response evaluation,[26] we also included a cohort of patients treated in the real-world setting with first-line chemotherapy, another cohort treated with cisplatin-gemcitabine and durvalumab and a final cohort of patients treated with second-line chemotherapy.

Protocol and registration

The protocol of the study was registered on the International Prospective Register of Systematic Reviews in February 2023 (PROSPERO registration ID CRD42023398279). Following a protocol amendment in October 2023, incorporating the patient-level data and an improvement in the search strategy, an updated systematic review and a new analysis were performed (see Protocol). We followed the PRISMA reporting guidelines.[27] The study was approved by the Vall d'Hebron Research Ethics Committee (PR(AG) 29/2024).

Search strategy

We searched Medline through Pubmed, Embase, Cochrane Library and ClinicalTrials.gov databases from inception to October 2023 (Table S1). Additionally, we searched references of the selected studies and abstract proceedings from the American Society of Clinical Oncology (ASCO), European Society of Medical Oncology (ESMO), ASCO Gastrointestinal Cancers Symposium (ASCO-GI), ESMO World Congress on Gastrointestinal Cancer and ESMO Asia.

The title and abstract of non-English studies were translated into English for the first screening step. The full text of those studies considered eligible for further evaluation was then translated. Of note, we identified no non-English study that required full-text evaluation.

All abstracts were reviewed and independently evaluated by two investigators through the Rayyan interface. Any disagreements were resolved by consensus with a third reviewer.

Eligibility criteria

Eligible studies were comparative phase II-III RCTs assessing systemic agents in the treatment of advanced BTC and included OS, PFS and/or ORR/DCR as an endpoint (Table S2). Studies that assessed locoregional or maintenance therapies, involved tumors other than BTC (except for periampullary carcinomas), were non-randomized, non-comparative or included patients in the (neo)adjuvant settings were excluded. The most recent and updated version of the trial was included in the final analysis.

Data extraction and quality assessment

We extracted the following data from the available reports: trial and baseline patient characteristics, number of patients included, endpoints, intervention details, median follow-up, response assessment criteria, OS hazard ratio (HR), PFS HR, ORR and DCR.

We generated funnel plots to assess publication bias (taking the 95% CIs to account for the heterogeneity estimated by the model) and used Egger's regression test to assess funnel plot asymmetry. Additionally, a *p* curve analysis was used to assess any further publication bias.

To assess the methodological quality of the included studies, we used two distinct tools: the Cochrane Handbook for Systematic Reviews of Interventions Risk of Bias tool (RoB version 2.0)[28] and the Delphi list.[29] Reports with a low or moderate risk of bias according to Cochrane's RoB or a score ≥5 points in the Delphi list were considered high quality.

A description of patients and datasets used for the individual-level correlation can be found in the supplementary materials and methods.

Statistical analyses

Condition 1 (Trial-level): All extracted endpoints were collected as defined by the trial. For trials that did not report HR, we estimated these with the methods described by Tierney *et al.* The odds ratio (OR) estimates for ORR and DCR were obtained from logistic regression models, including patients with measurable disease and considering non-evaluable patients as non-responders. The HR and OR were log-transformed and the associations estimated using a linear regression model weighed by trial size. The variation of the weighted treatment effects explained by the model was measured with the coefficient of determination (R^2).

The surrogate threshold effect (STE) represents the minimum treatment effect of the intermediate endpoint needed to predict a non-zero effect on OS and is calculated based on the prediction interval. The 95% prediction intervals were constructed for the regression line of the treatment effect on OS *vs*. the surrogate with a weight (*i.e*. trial size) of 200. The STE was defined as the intersection of the upper 95% prediction interval with the horizontal y-axis = 0, representing a hazard ratio of 1.[30,31]

We analyzed predefined subgroups according to the presence of crossover, trial size, type of treatment, disease setting/line of treatment and quality of the trials. We further performed two non-preplanned sensitivity analyses based on disease location and stage by assigning each trial a weight proportional to the number of included patients for each category. Additionally, we performed a leave-one-out cross-validation, whereby each trial was left out once, and the model was refitted with the remaining trials. The resulting model was then applied to the left-out trial to predict the effect of treatment on the reference endpoints. The R^2 of the cross-validated model was calculated as the correlation between the individual predictions made by the model and the actual treatment effects.[32]

Condition 2 (Patient-level): The correlation between OS and PFS was measured by using the normal score rank correlation, calculated by the iterative multiple imputation approach.[33] Although this approach is semiparametric and does not require any assumptions about the marginal distributions, it uses a Gaussian dependency structure. Therefore, we also calculated the rank correlation between OS and PFS using a non-parametric estimator of Spearman's correlation, based on a non-parametric bivariate survival surface estimator.[34] The 95% CIs were calculated by bootstrap resampling 1,000 times.

To evaluate the association between response and OS, we performed a responder analysis.[35–37] Responders were defined as patients who achieved a partial or complete response and non-responders as those with stable disease, progressive disease or whose response status was unknown or non-evaluable. To adjust for immortal-time bias, a landmark analysis was performed[38] at 3-month and 6-month landmark times for first-line trials and 2-month and 4-month times for second-line trials. Only the datasets of patients included in randomized trials were used for this analysis, as no longitudinal response assessment was available for the RWD cohorts.

We scored the strength following the criteria used by Prasad *et al.*:[39] low correlation ($r \leq 0.7$), moderate strength correlation ($r > 0.7$ to $r < 0.85$), and high correlation ($r \geq 0.85$).

All statistical analyses were completed using R version 4.1.2 (R Foundation).

Results

Condition 1: Trial-level association

Of the 8,576 records identified, a total of 41 randomized phase II and phase III clinical trials were eligible, including 44 treatment comparisons, 88 treatment arms and 7,817 patients (Fig. 1; Tables 1, 2 and S3). Most studies were phase II trials (70.7%), included first-line combinations (65.9%), tested chemotherapy (53.7%) or targeted/tyrosine kinase inhibitor (36.6%) agents, and were multicenter (80.5%), while only 2 (4.9%) allowed for crossover. The median follow-up was 10.85 months (IQR 10.1-15.7 months), although 17 (41.5%) trials did not report this information. Twenty trials (48.8%) used PFS as a primary endpoint and 16 (39%) used OS. Eleven (26.8%) were double blind and the remaining 30 trials were open-label.

We found no evidence of publication bias by applying the two distinct detection methods for OS, PFS and ORR (Fig. S1). A funnel plot asymmetry was detected for DCR, although the *p*-curve analysis showed that evidential value was present. The overall risk of bias was low or moderate, and only two studies were found to be at high risk of bias (Fig. S2 and S3). When applying the Delphi assessment criteria,[29] 35 studies were found to be of high quality, and six had a score below 5 points (Fig. S4).

The correlation between PFS and OS showed an R^2 of 0.71 (95% CI 0.56-0.86) and the STE was 0.61 (Fig. 2A), meaning that a HR of 0.61 in a hypothetical trial of 200 patients would likely lead to a non-zero effect on OS. Importantly, the correlations with ORR and DCR were low or non-existent, with R^2 values of 0.01 (95% CI 0-0.08) and 0.39 (95% CI 0.14-0.64), respectively (Fig. 2B,C). Prespecified subgroup analyses based on the line of treatment, presence of crossover, study phase, type of systemic treatment, sample size and trial quality confirmed these findings (Figs 2D, 3, S5 and S6). Non-preplanned sensitivity analyses showed consistent results for distinct disease locations and stages (Fig. S7–S9). The correlation of ORR and DCR with OS remained low across all subgroups. We further calculated the STE for all surrogate endpoints based on different hypothetical sample sizes (Table S4).

Finally, we performed a leave-one-out cross-validation procedure to confirm the correlation observed between OS and PFS. The R^2 ranged from 0.61 to 0.78. All trial HR estimates for OS fell within the predicted intervals except for three (Fig. S10). Two of these were highly influential trials in the cross-validation: the ClarIDHy trial,[40] whose exclusion from the model led to an R^2 of 0.78, and the NuTide:121,[41] whose exclusion led to an R^2 of 0.61. The R^2 remained consistent after individually excluding the remaining trials, with R^2 values that ranged from 0.7 to 0.73 (Fig. S10B).

Condition 2: Patient-level association

We analyzed five datasets involving 2,506 patients diagnosed with advanced BTC who received systemic treatments: a pooled population of 512 patients included in the ABC-01,[20] -02[21] and -03[22] trials, a RWD dataset of 628 patients treated with first-line cisplatin-gemcitabine and durvalumab, a RWD dataset of 773 patients treated with first-line chemotherapy, a pooled population of 271 patients included in the NIFTY[23] and

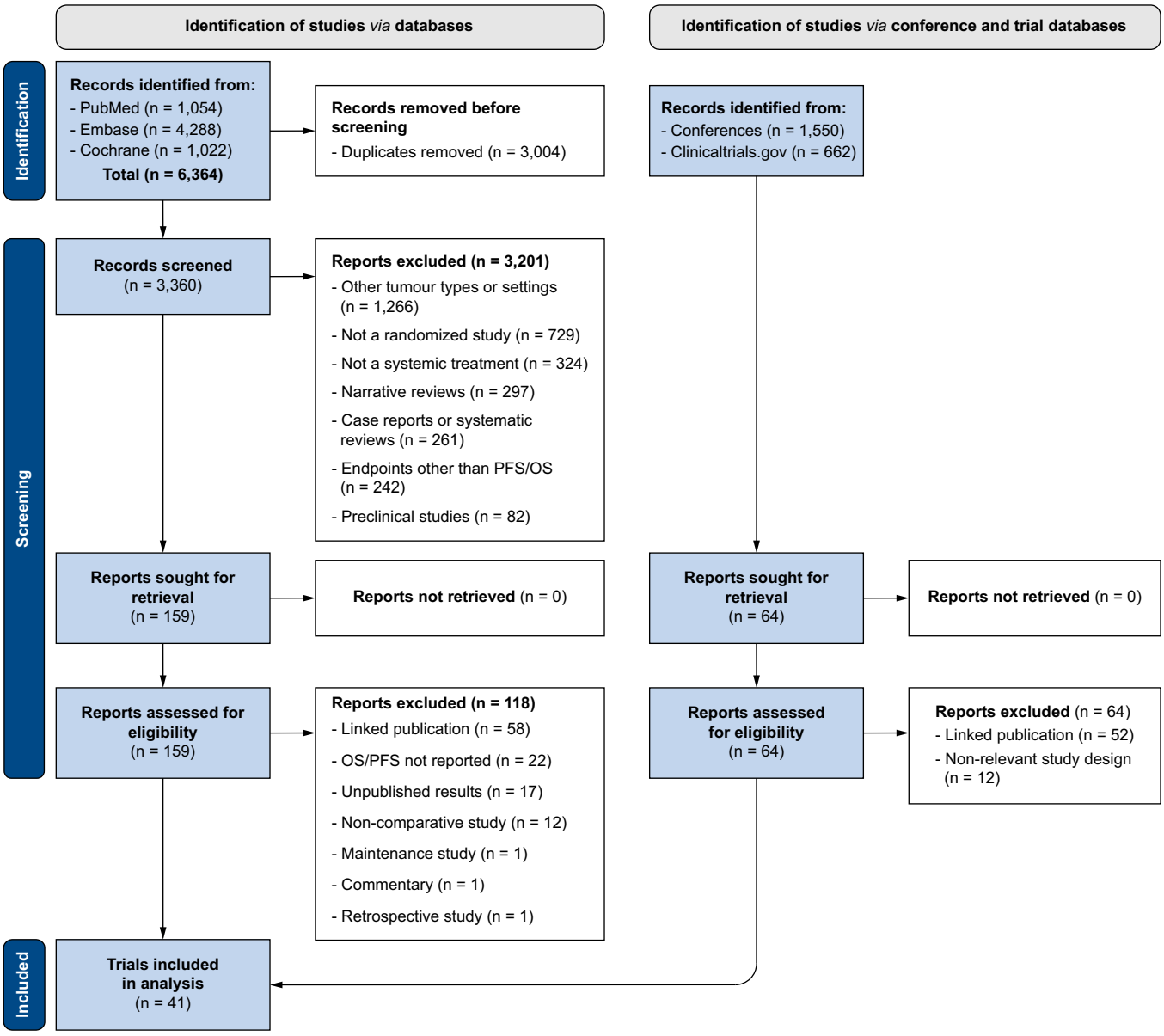


Fig. 1. PRISMA flowchart reporting the results of the systematic review.

FIReFOX[24] trials and a RWD dataset of 322 patients treated with standard second-line chemotherapy (Table S5–S8).

We estimated the correlation between PFS and OS at the patient-level using two distinct methods. We applied the multiple imputation approach[33] and found a rank correlation ranging between 0.73 and 0.86 across all five datasets (Table 3). Only the pooled population of NIFTY[23] and FIReFOX[24] trials showed a slightly lower correlation of 0.73, while all other datasets showed a rank correlation above 0.8. We also calculated the correlation using a more conservative, non-parametric estimator of Spearman's correlation.[34] This approach rendered similar results, although the correlation estimated by this method tended to be lower in all datasets, with a rank correlation that ranged between 0.68 and 0.82 (Table S9). We found consistent results across distinct disease locations and stages (Table S10 and S11).

Finally, to estimate the association between ORR and OS, we performed a responder analysis.[35–37] We only included datasets of patients treated in RCTs, as longitudinal response data was not available in the RWD. In the first-line setting, 370 patients had measurable disease and were included in this analysis. The ORR was 23% and the DCR was 77.8%. Responders did not experience better survival, either at the 3-month or 6-month landmark times (Fig. 4). In the second-line setting, 256 patients were included in the analysis. The ORR was 9.4% and the DCR was 64.1%. Similar to the first-line setting, responders did not experience better survival (Fig. S11). However, given the low response rate in the second-

Table 1. Characteristics of the trials, treatment comparisons and patients included in the studies.

	Patients (n = 7,817)	Trials (n = 41)	Comparisons (n = 44)[1]
Age, median (IQR)	63 (60.5–64)	-	-
Missing	151 (1.9%)		
Sex, n (%)		-	-
Males	3,952 (50.6%)		
Females	3,818 (48.8%)		
Missing	47 (0.6%)		
Tumor location, n (%)			
Intrahepatic	3,445 (44.1%)	37 (90.2%)	38 (86.4%)
Extrahepatic	1,410 (18%)	36 (87.8%)	37 (84.1%)
Gallbladder	2,272 (29.1%)	40 (97.6%)	43 (97.7%)
Ampullary	192 (2.5%)	17 (41.5%)	18 (40.9%)
Other	36 (0.5%)	-	-
Missing	462 (5.9%)	-	-
Disease stage, n (%)			
Locally advanced	1,457 (18.6%)	38[2] (95%)	41[2] (95.3%)
Metastatic	5,810 (74.3%)	41 (100%)	44 (100%)
Missing	550 (7%)	-	-
ECOG status, n (%)			
0	3,194 (40.9%)	40 (97.6%)	42 (95.5%)
1	3,569 (45.7%)	40 (97.6%)	42 (95.5%)
2-3	311 (4%)	14 (34.1%)	16 (36.4%)
Missing	743 (9.5%)	-	-
Number of centers, n (%)			
Multicenter	7,088 (90.7%)	33 (80.5%)	34 (77.3%)
Single-center	634 (8.1%)	6 (14.6%)	8 (18.2%)
Missing	95 (1.2%)	2 (4.9%)	2 (4.5%)
Treatment line, n (%)			
First line	6,164 (78.9%)	27 (65.9%)	29 (65.9%)
Beyond first line	1,653 (21.1%)	14 (34.1%)	15 (34.1%)
Systemic agents, n (%)			
Chemotherapy	5,295 (67.7%)	22[3] (53.7%)	23 (52.3%)
Immunotherapy	1,153 (14.7%)	5 (12.2%)	5 (11.4%)
Targeted therapy	1,066 (13.6%)	15[3] (36.6%)	16 (36.4%)
Placebo/BSC	303 (3.9%)	-	-
Clinical trial phase, n (%)			
Phase II	2,814 (36%)	29 (70.7%)	31 (70.5%)
Phase III	5,003 (64%)	12 (29.3%)	13 (29.5%)
Crossover, n (%)	53 (0.7%)	2 (4.9%)	2 (4.5%)
Follow-up (months), median (IQR)	-	10.85 (10.1–15.7)	-
Missing		17 (41.5%)	

BSC, best supportive care; ECOG, Eastern Cooperative Oncology Group.
[1]Three trials[54–56] contained three arms, leading to two comparisons.
[2]One trial[57] did not specify whether locally advanced patients were included.
[3]One trial[55] had two experimental arms, one including chemotherapy and another targeted therapy.

line setting, the number of responders at each landmark time is low and precludes any definitive conclusions.

Discussion

Inappropriate validation of intermediate endpoints may lead to the approval of potentially ineffective or even harmful treatments.[10] Previous studies have evaluated the association of PFS and response with OS at the trial-level in the context of first-[12,14] and second-line treatment of BTC.[11,13] Our analysis can be distinguished from these studies in several important ways: First, it is the only one, to our knowledge, that has included trial- and patient-level data. Second, it includes contemporary trials testing distinct systemic agents, including immunotherapies and targeted agents. Third, we rigorously extracted and calculated the HR for time-to-event endpoints and OR for binomial endpoints to ensure homogeneous analyses of these variables and appropriate measures of treatment effect. Finally, we also included RWD to complement the RCT information.

This comprehensive analysis suggests that the correlation for PFS is moderate both at the trial- and patient-level but is low for ORR and DCR in advanced BTC. Whether the strength of the correlation is sufficient to justify the use of PFS as a surrogate endpoint is arguable and controversial. For instance, PFS would meet the surrogacy criteria established by the BSES[42,43] and ReSEEM[17] guidelines, while the IQWiG[44] guidelines would consider the evidence "Unclear". Regulatory agencies have not established criteria for defining surrogate endpoints. We believe that PFS could be used as a primary endpoint in advanced BTC in circumstances when OS may be confounded, such as crossover designs or accelerated approval programs that may lead to uncontrolled post-progression crossover in the confirmatory trial. In these circumstances, a careful evaluation of OS should continue to be mandatory to ensure no detrimental effect is observed.[5,45] The magnitude of the benefit in PFS should also be considered. Our analysis of the STE shows that a magnitude of 0.61 (0.67 after excluding crossover trials) in PFS would likely lead to an OS

Table 2. Characteristics and design of the trials included in the systematic review.

Trial[1]	Treatments	Phase	N	Blinding	Primary Endpoint	Response evaluation	Timing of scans
ABC-02	CG Gemcitabine	III	410	Open-label	OS	RECIST 1.0	Q12w
ABC-03	CG+cediranib CG+placebo	II	124	Double blind	PFS	RECIST 1.1	Q12w
BilT-01	Nivo-ipi CG-nivo	II	68	Open-label	PFS 6 months	RECIST 1.1/irRECIST	Q8w
BREGO	mGEMOX+regorafenib mGEMOX	II	66	Open-label	NA	RECIST 1.0	NA
BT22	CG Gemcitabine	II	83	Open-label	OS 1 year	NA	Q8w
Chen 2015	GEMOX+cetuximab GEMOX	II	122	Open-label	ORR	RECIST 1.1	Q8w
ClarIDHy	Ivosidenib Placebo	III	187	Double blind	PFS	RECIST 1.1	Q6w
FIReFOX	mFOLFIRI mFOLFOX	II	118	Open-label	OS 6 months	RECIST 1.1	Q6w
Gambit	Irinotecan+Cisplatin CG	II	47	Open-label	ORR	RECIST 1.1	NA
GB-SELECT	CAPIRI Irinotecan	II	98	Open-label	OS 6 months	RECIST 1.1	Q8w
GEMSO-AIO	Gemcitabine+sorafenib Gemcitabine	II	97	Double blind	PFS	RECIST 1.0	Q8w
Ikeda 2023	Nanvuranlat Placebo	II	104	Double blind	PFS	RECIST 1.1	NA
IMbrave151	CG+atezolizumab+bevacizumab CG+atezolizumab+placebo	II	162	Double blind	PFS	RECIST 1.1	Q9w
JCOG0805	SG S1	II	101	Open-label	OS 1 year	RECIST 1.0	Q6w
JCOG1113	SG CG	III	354	Open-label	OS	RECIST 1.1	Q6w
Kang 2012	SG CG	II	96	Open-label	PFS 6 months	RECIST 1.0	Q6w
Kataria 2022	Capecitabine BSC	II/III	69	Open-label	OS	RECIST 1.1	NA
Kataria 2022	Erlotinib BSC	II/III	69	Open-label	OS	RECIST 1.1	NA
KEYNOTE-966	CG+pembrolizumab CG-placebo	III	1,069	Double blind	OS	RECIST 1.1	Q6w
KHBO1401-MITSUBA	CG CGS	III	246	Open-label	OS	RECIST 1.1	Q12w
Kim 2019	CAPOX GEMOX	III	222	Open-label	PFS	RECIST 1.1	Q6w
Lee 2012	GEMOX+erlotinib GEMOX	III	268	Open-label	PFS	RECIST 1.0	Q6w
Markussen 2020	GEMOX-capecitabine CG	II	96	Open-label	PFS	RECIST 1.1	Q12w
NALIRICC	5FU-nalIRI 5FU	II	100	Open-label	PFS	RECIST 1.1	Q6w
NIFTY	5FU-nalIRI 5FU	II	174	Open-label	PFS	RECIST 1.1	Q6w

(*continued on next page*)

Table 2. (*continued*)

Trial[1]	Treatments	Phase	N	Blinding	Primary Endpoint	Response evaluation	Timing of scans
Nutide:121	Cisplatin+NUC1031 CG	III	773	Open-label	OS, ORR	RECIST 1.1	Q9w
Pape 2020	CAP7.1 BSC	II	27	Open-label	DCR	RECIST 1.1	Q8w
PICCA	CG+panitumumab CG	II	90	Open-label	PFS 6 months	RECIST 1.0	Q6w
REACHIN	Regorafenib Placebo	II	66	Double blind	PFS	RECIST 1.1	Q6w
Schinzari 2017	FOLFOX4 De Gramont	II	48	Open-label	OS	RECIST 1.1	Q8w
Sharma 2010	mGEMOX BSC	II	53	Open-label	OS, ORR, toxicity	RECIST 1.0	Q6w
Sharma 2010	FUFA BSC	II	55	Open-label	OS, ORR, toxicity	RECIST 1.0	Q6w
Sharma 2019	mGEMOX CG	III	243	Open-label	OS	RECIST 1.1	NA
Shirahama 2017	PPV+CPA PPV	II	49	Open-label	Immune response	RECIST 1.0	Q8w
SWOG 1815	CG+Nab/paclitaxel CG	III	441	Open-label	OS	RECIST 1.1	Q9w
SWOG S1310	Trametinib 5FU/capecitabine	II	44	Open-label	OS	RECIST 1.1	Q6w
TOPAZ-1	CG+durvalumab CG	III	685	Double blind	OS	RECIST 1.1	Q6w
TreeTopp	Varlitinib+capecitabine Placebo+capecitabine	II	127	Double blind	ORR, PFS	RECIST 1.1	Q6w
Ueno 2021	Reminostat+S1 Placebo+S1	II	101	Double blind	PFS	RECIST 1.1	Q6w
Valle 2021	Ramucirumab Placebo	II	207	Double blind	PFS	RECIST 1.1	Q6w
Valle 2021	Merestinib Placebo	II	203	Double blind	PFS	RECIST 1.1	Q6w
Vecti-BIL	GEMOX+panitumumab GEMOX	II	89	Open-label	PFS	RECIST 1.1	Q8w
Yang 2022	Cisplatin+Nab/paclitaxel CG	II	67	Open-label	PFS	NA	NA
Zheng 2018	XELIRI Irinotecan	II	60	Open-label	PFS	RECIST 1.1	Q6w

BSC, best supportive care; CG, cisplatin + gemcitabine; DCR, disease control rate; NA, not available; ORR, objective response rate; OS, overall survival; PFS, progression-free survival.
[1]A complete list of the studies referenced in the table is found in the Supplementary Materials.

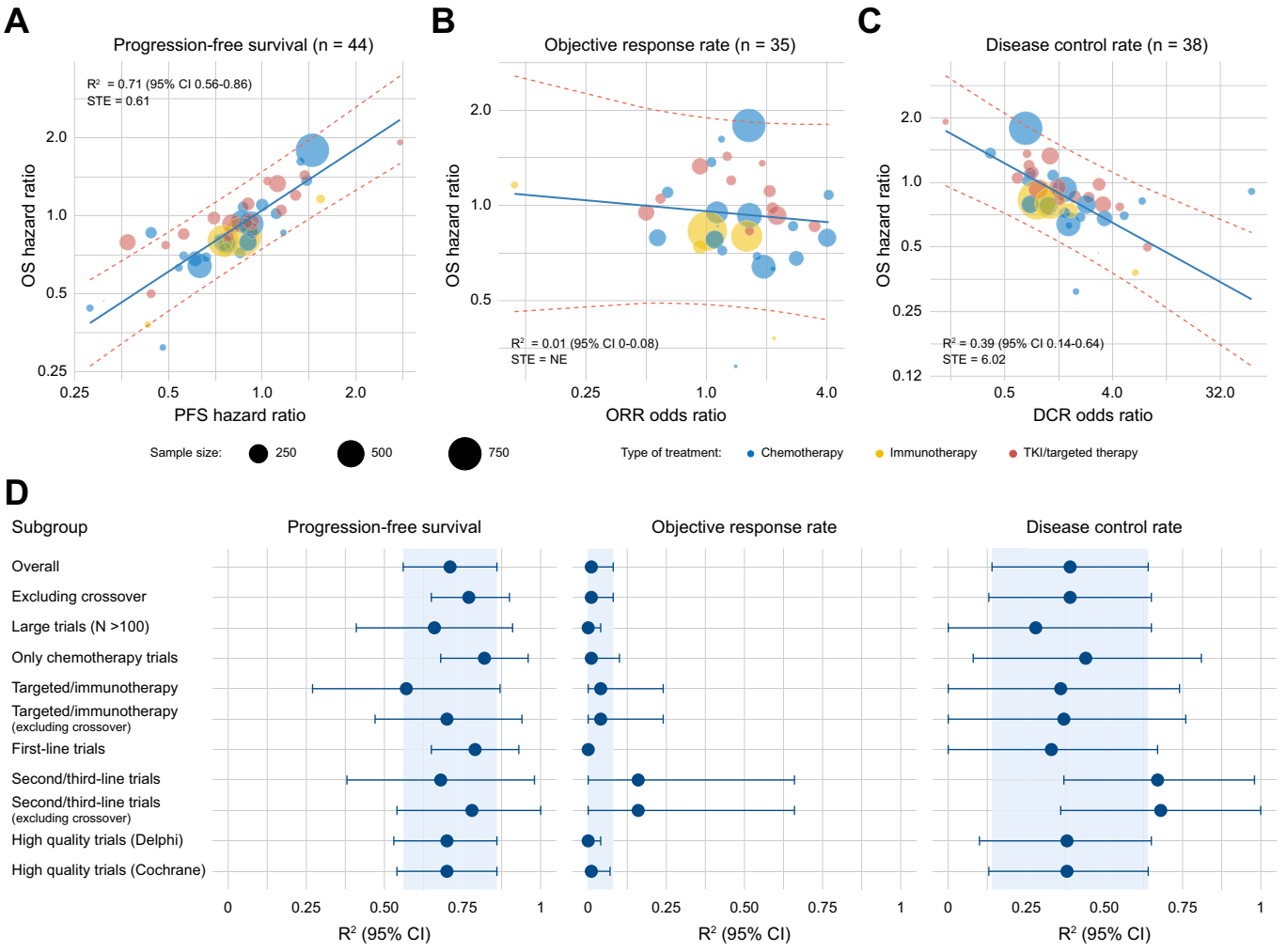


Fig. 2. Trial-level correlation of PFS, ORR and DCR with OS (Condition 1). Bubble plot assessing the correlation of (A) PFS, (B) ORR and (C) DCR with OS. Every bubble represents a trial, the color represents the treatment type and the size is proportional to the number of patients included in the trial. The hazard ratios and odds ratios are presented in the logarithmic scale. The red lines show the 95% prediction interval for a weight (*i.e.* sample size) of 200. (D) Subgroup analysis assessing the coefficient of determination across different subgroups according to the indicated criteria (left column). The shaded blue rectangles indicate the reference R^2 and its 95% CI of the global correlation. The correlation was estimated by using a linear regression model weighed by trial size. The variation of the weighted treatment effects explained by the model was measured with the coefficient of determination (R^2). DCR, disease control rate; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; STE, surrogate threshold effect. (This figure appears in color on the web.)

benefit in a 200-patient randomized trial, which may be informative for interpreting and designing future studies. In addition, the use of a STE may provide a more reasonable standard for evaluating the magnitude of a treatment benefit in BTC when using PFS as a surrogate endpoint, such as the ones proposed by the ESMO Magnitude of Clinical Benefit Scale guidelines.[46] However, our data also highlight the importance of further refining and developing novel endpoints. In the case of PFS, for example, the use of time to treatment failure, which incorporates treatment discontinuation due to toxicity as an event to avoid informative censoring[26,47–49] or considering the pattern of progression may help to better capture OS.[50]

The results of our study do not support the use of either ORR or DCR as surrogate endpoints in this setting. Several factors may account for this finding. First, BTCs are frequently infiltrative and irregular, making it challenging to radiologically monitor the disease.[51] Second, patients who do not achieve a response might not be uniformly disadvantaged, especially when receiving non-cytotoxic agents, as these may confer improved survival by restraining tumor progression without inducing radiological responses.[3,40] Third, BTCs are densely fibrotic tumors in which treatment-induced tumor death may not necessarily lead to tumor shrinkage. Other parameters, such as metabolic changes, may be more accurate in discriminating response.[52] Finally, the low ORR observed with most systemic therapies in BTC may decrease the prognostic discrimination of response and lead to this poor correlation.

Several limitations should be considered when interpreting this study. First, the systematic review included a heterogeneous group of trials involving different study lines, treatment regimens and patient populations. Nonetheless, this high heterogeneity is necessary to support the assertion of the validity of a surrogate for application in a new trial.[53] Additionally, we conducted several predefined subgroup and sensitivity analyses which showed consistent levels of correlation. Second,

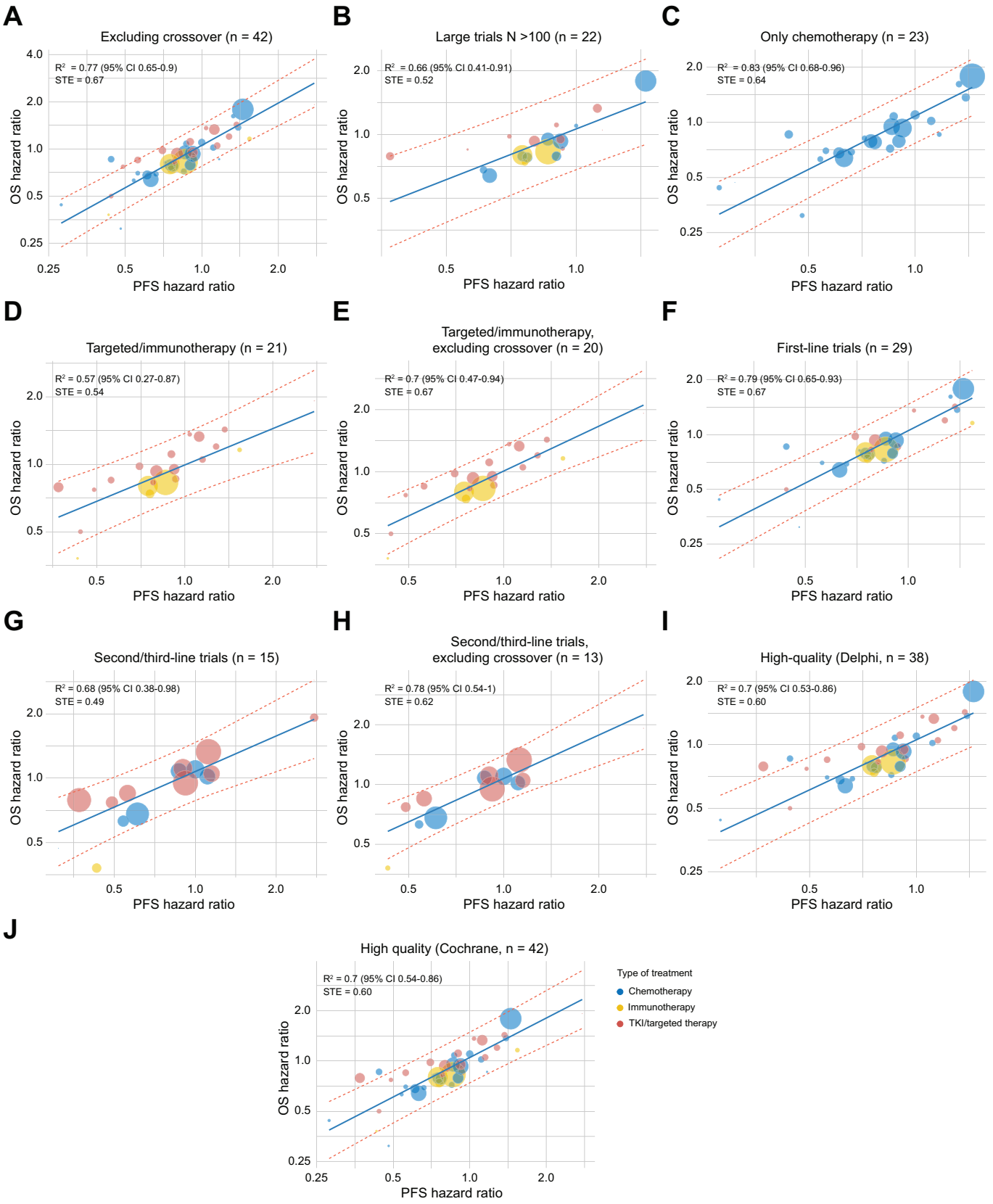


Fig. 3. Bubble plots showing the trial-level correlation for PFS and OS across different subgroups. Every bubble represents a trial, the color represents the treatment type and the size is proportional to the number of patients included in the trial. The hazard ratios are presented in the logarithmic scale. The red lines show the 95% prediction interval for a weight (*i.e.* sample size) of 200. The correlation was estimated by using a linear regression model weighed by trial size. The variation of the weighted treatment effects explained by the model was measured with the coefficient of determination (R^2). OS, overall survival; PFS, progression-free survival; STE, surrogate threshold effect. (This figure appears in color on the web.)

Table 3. Patient-level correlation of PFS with OS across the different datasets using the iterative imputation method.

Cohort	Setting	Treatment line	N (events)	Follow-up (95% CI)	Median OS (95% CI)	Median PFS (95% CI)	ρ_{imi} (95% CI)
ABC-01, -02, -03	RCT	First line	512 (497)	51 (41.1-NA)	10.2 (9-11.5)	6.5 (6-7.4)	0.84 (0.81-0.86)
DURVABTC	RWD	First line	628 (190)	8.4 (7.8-9.4)	14.9 (13.4-17.8)	8.2 (7.5-8.9)	0.86 (0.81-0.9)
RETUD	RWD	First line	773 (623)	32 (25.3-37.3)	9.7 (8.7-10.4)	5 (4.5-5.4)	0.83 (0.8-0.85)
NIFTY, FIReFOX	RCT	Second line	277 (236)	33 (27-37.2)	6.3 (5.5-7.4)	2.6 (2.4-2.9)	0.73 (0.67-0.79)
RETUD	RWD	Second line	322 (279)	24.8 (22.3-NA)	5.2 (4.8-6)	2.8 (2.5-3)	0.81 (0.78-0.83)

The correlation coefficient ρ_{imi} was measured by using a normal score rank correlation calculated by the iterative multiple imputation approach.
Follow-up, OS, and PFS measured in months.
NA, not available; OS, overall survival; PFS, progression-free survival; RCT, randomized-controlled trial; RWD, real-world data.

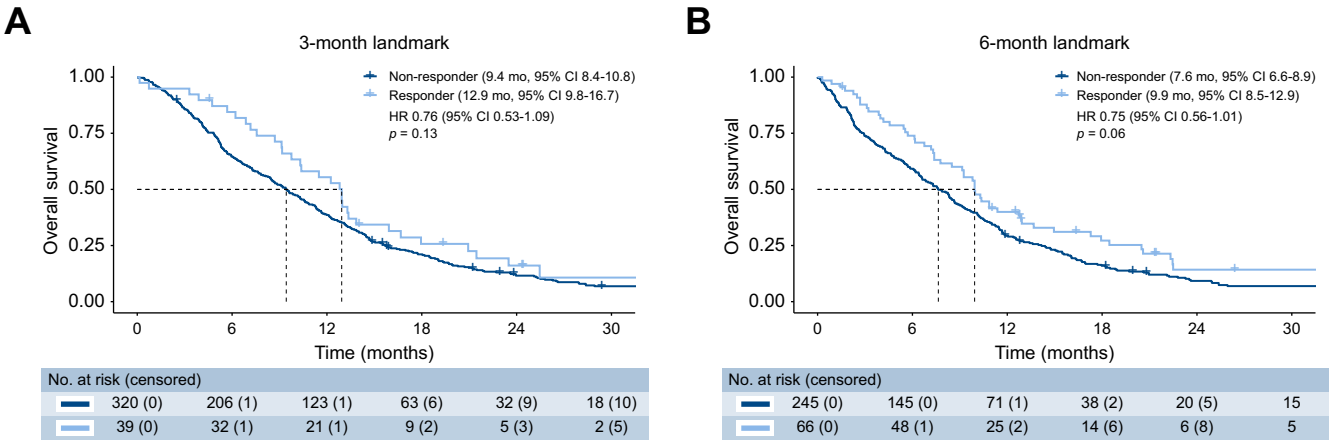


Fig. 4. Impact of response on survival in patients treated with first-line chemotherapy. Kaplan-Meier estimates of overall survival between responders and non-responders (Condition 2) who were alive and had achieved response at (A) 3-month and (B) 6-month landmark times. The HRs were estimated by applying a Cox regression model and the *p* values obtained from the Cox regression model. HR, hazard ratios.

the trial-level correlation was performed with aggregate data rather than patient-level data. We intentionally modelled the trial-level and individual-level correlation separately to ensure a broad inclusion of trials in the first condition and decrease the risk of selection bias. Third, most of the trials explored chemotherapy or tyrosine kinase inhibitors. The meta-analysis will have to be updated when further randomized studies exploring immunotherapy combinations and targeted therapies (especially FGFR inhibitors) become available. Importantly, the association of ORR/DCR and PFS with OS will have to be confirmed in individual-level data for patients treated with these therapies. Fourth, trials did not uniformly time the radiological assessments nor use a uniform definition for response evaluation. Although this may influence PFS and ORR/DCR, it is reflective of the current scenario of RCTs and highlights the need to establish a uniform set of criteria for defining and evaluating PFS in future trials.

In conclusion, our results caution against the routine use of surrogate endpoints in randomized trials testing systemic agents in advanced BTC and highlight the need for further developments to better capture OS. However, until better surrogate endpoints are developed and validated, PFS should be prioritized over ORR and DCR. Furthermore, validation in RCTs including targeted therapies and immunotherapies will be necessary to confidently extrapolate these results to trials assessing these therapies.

Affiliations
[1]*Upper Gastrointestinal and Endocrine Tumor Unit, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital, Barcelona, Spain;* [2]*Catalan Institute of Oncology (ICO), L'Hospitalet de Llobregat, Barcelona, Spain;* [3]*Department of Medical Oncology, University College London Cancer Institute, London, United Kingdom;* [4]*Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea;* [5]*Department of Oncology, Vita-Salute San Raffaele University, IRCCS San Raffaele Scientific Institute Hospital, Milan, Italy;* [6]*Department of Medical Oncology, UGC Oncología Guipúzcoa, País Vasco, Spain;* [7]*Department of Medical Oncology - Onco Health Institute, Instituto de Investigaciones Sanitarias FJD, Fundación Jiménez Díaz University Hospital, Madrid, Spain;* [8]*Department of Medical Oncology, The Christie NHS Foundation, Manchester, Division of Cancer Sciences, University of Manchester, Manchester, United Kingdom;* [9]*Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy and Sant'Anna School of Advanced Studies, Pisa, Italy;* [10]*Department of Medical Oncology, Asturias Central University Hospital, ISPA, Oviedo, Spain;* [11]*Cancer Research UK & UCL Cancer Centre, University College of London, London, United Kingdom;* [12]*Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea;* [13]*Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (Milan), Italy;* [14]*Medical Oncology and Hematology Unit, Humanitas Cancer Center, IRCCS Humanitas Research Hospital, Rozzano (Milan), Italy;* [15]*Department of Medical Oncology Hospital Universitario 12 de Octubre, Madrid, Spain;* [16]*Biblioteca, Docència, Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain;* [17]*Medical Oncology Service, Hospital General Universitario Gregorio Marañón, Universidad Complutense, Madrid, Spain;* [18]*Cholangiocarcinoma Foundation, Herriman, UT, United States;* [19]*Division of Cancer Sciences, University of Manchester, Manchester, United Kingdom*

Abbreviations

ASCO, American Society of Clinical Oncology; BTC, biliary tract cancer; DCR, disease control rate; ESMO, European Society of Medical Oncology; HR, hazard ratio; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; RCT, randomized-controlled trial; RWD, real-world data; STE, surrogate threshold effect.

Financial support

The study received no financial support.

Conflict of interest

F Castet: Received honoraria from AstraZeneca, Eisai, Roche, Servier; and travel expenses from Roche and Servier. C Fabregat-Franco: None. J Bridgewater: None. JW Kim: Received advisory and consulting fees from AstraZeneca, BeiGene, Beyond Bio, Bristol Myers Squibb, Celgene, Eisai, GC Cell, MSD, ONO, Sanofi-Aventis, Servier, TCUBEit; and research funding paid to his institution from Samyang biopharm, Boryung. M Rimini: None. A La Casta: None. A Lamarca: received personal fees from NETConnect Initiatives funded by Ipsen; honoraria from Merck, Pfizer, Ipsen, Incyte, AAA/Novartis, QED, Servier, Astra Zeneca, EISAI, Roche, Advanz Pharma and MSD; advisory or consulting fees from EISAI, Nutricia, Ipsen, QED, Roche, Servier, Boston Scientific, Albireo Pharma, AstraZeneca, Boehringer Ingelheim, GENFIT, TransThera Biosciences, Taiho and MSD; travel expenses from Ipsen, Pfizer, Bayer, AAA, SirtEx, Novartis, Mylan, Delcath Advanz Pharma and Roche; research funding paid to her institution from QED, Merck, Boehringer Ingelheim, Servier, AstraZeneca, GenFit, Panbela Therapeutics, Novocure GmbH, Camurus AB, Albireo Pharma, Taiho, TransThera, Jazz Therapeutics and Roche. M Kang: None. F Salani: Received honoraria from Daiichi Sankyo; and travel expenses from Leo Pharma. A Castillo: None. A Lopes: None. J Hyung: None. L Rimassa: Received honoraria from AstraZeneca, Bayer, BMS, Guerbet, Incyte, Ipsen, Roche, Servier; advisory or consulting fees from AbbVie, AstraZeneca, Basilea, Bayer, BMS, Elevar Therapeutics, Exelixis, Genenta, Hengrui, Incyte, Ipsen, IQVIA, Jazz Pharmaceuticals, MSD, Nerviano Medical Sciences, Roche, Servier, Taiho Oncology, Zymeworks; travel expenses from AstraZeneca; research funding to the institution from Agios, AstraZeneca, BeiGene, Eisai, Exelixis, Fibrogen, Incyte, Ipsen, Lilly, MSD, Nerviano Medical Sciences, Roche, Servier, Taiho Oncology, TransThera Sciences, Zymeworks. J Adeva: Received honoraria from AstraZeneca, Roche, Servier; advisory or consulting fees from AstraZeneca, Jazz Pharmaceuticals, MSD, Roche, Servier, Taiho Oncology, Zymeworks; and travel expenses from AstraZeneca, Roche, Servier. D López-Valbuena: None. M Basagaña-Farrés: None. S Vaja: None. KM Mak: None. TV Tian: Received honoraria from AstraZeneca, Incyte, and Jazz Pharmaceuticals; research funding to his instution from AstraZeneca, LOXO Oncology, Servier, Alentis, and Incyte. A Muñoz: Received personal fees from Rovi, Menarini, Stada, Medscape; advisory or consulting fees from GSK, Pfizer, BMS-Celgene, Sanofi, Astra-Zeneca, MSD, Lilly, Servier, Roche, Taiho, Leo Pharma; research funding to his institution from Leo Pharma, Sanofi, Celgene; patent for Risk assessment model in venous thromboembolism in cancer patients. A Casadei-Gardini: Received personal fees from AstraZeneca, Bayer, BMS, Eisai, Incyte, Ipsen, Roche, Servier; advisory or consulting fees from AstraZeneca, Bayer, BMS, Eisai, Incyte, Ipsen, IQVIA, MSD, Roche, Servier; travel expenses from AstraZeneca; and research funding to institution from AstraZeneca, Eisai. C Yoo: Received honoraria from Servier, Bayer, AstraZeneca, Merck Sharp & Dohme, Eisai, Celgene, Bristol Myers Squibb, Ipsen, Novartis, Boryung, Mundipharma, Roche; research funding to institution from Servier, Bayer, AstraZeneca, Ono Pharmaceuticals, Ipsen, Boryung, Lunit Inc. JW Valle: None. T Macarulla: Received honoraria from Janssen, Lilly, Esteve, Daïchi, Biontech, Novartis, Jazz Pharmaceuticals; advisory or consulting fees from Ability Pharmaceuticals SL, Arcus Bioscience Inc., AstraZeneca, Basilea Pharma, Baxter, BioLineRX Ltd, Celgene, Eisai, Incyte, Ipsen Bioscience Inc; travel expenses from Servier, AstraZeneca, Sanofi, Incyte, Lilly, MSD and Roche; research funding to institution MSD, Novocure, QED Therapeutics, Roche Farma, Sanofi-Aventis, Servier, Zymeworks.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

FC: acquisition of data; data analysis and interpretation; manuscript writing, study concept and design. ***CFF***: acquisition of data; interpretation of data; critical revision of the manuscript. ***JB***: acquisition of data; interpretation of data; critical revision of the manuscript. ***JWK***: acquisition of data; interpretation of data; critical revision of the manuscript. ***MR***: acquisition of data; interpretation of data; critical revision of the manuscript. ***ALC***: acquisition of data; interpretation of data; critical revision of the manuscript. ***AL***: acquisition of data; interpretation of data; critical revision of the manuscript. ***MK***: acquisition of data; interpretation of data; critical revision of the manuscript. ***FS***: acquisition of data; interpretation of data; critical revision of the manuscript. ***AC***: acquisition of data; interpretation of data; critical revision of the manuscript. ***AL***: acquisition of data; interpretation of data; critical revision of the manuscript. ***JH***: acquisition of data; interpretation of data; critical revision of the manuscript. ***LR***: acquisition of data; interpretation of data; critical revision of the manuscript. ***JA***: acquisition of data; interpretation of data; critical revision of the manuscript. ***DLV***: acquisition of data; interpretation of data; critical revision of the manuscript. ***MBF***: acquisition of data; critical revision of the manuscript. ***SV***: acquisition of data; data analysis and interpretation; critical revision of the manuscript. ***KMK***: acquisition of data; critical revision of the manuscript. ***TVT***: data analysis and interpretation; critical revision of the manuscript. ***AM***: acquisition of data; interpretation of data; critical revision of the manuscript. ***ACG***: acquisition of data; interpretation of data; critical revision of the manuscript. ***CY***: acquisition of data; interpretation of data; critical revision of the manuscript. ***JWV***: acquisition of data; interpretation of data; critical revision of the manuscript. ***TM*** acquisition of data; data analysis and interpretation; manuscript writing, study concept and design.

Data availability

All trial-level data are presented in the Article or Supplementary Materials. Study protocol is available on PROSPERO, registration number CRD42023398279. Participant-level information cannot be shared due to confidentiality agreements. Requests for raw, individual-level data should be directed to the study Sponsors. All the codes used in the analysis can be provided to qualified researchers upon reasonable request to the corresponding author.

Acknowledgements

The authors thank Inma Ruiz de Mena for her administrative support in the transfer of the RETUD data, Cancer Research UK & UCL Cancer Trials Centre (UCL CTC) for sharing pseudonymised data of ABC-02 and ABC-03 and Cancer Research UK for supporting the original studies (ABC-02 and ABC-03).

Spanish Cooperative Group for the Treatment of Digestive Tumors (TTD)

Adelaida Lacasta[1], Alfredo Castillo[2], Jorge Adeva[3], Andrés Muñoz[4], Rosario Vidal-Tocino[5], Florian Castet[6], Enrique Aranda[7], Míriam Lobo[8], Paloma Peinado[9], Javier Sastre[10], Ana Fernández-Montes[11], Begoña Graña[12], Eva Martínez de Castro[13], Javier Gallego[14], Ruth Vera[15], Inmaculada Alés[16], Inmaculada Gallego[17], Teresa García García[18], Ismael Ghanem[19], Inma Ruiz de Mena[20], Berta Laquente[21], David Páez[22], Raquel Molina[23], Mercedes Rodríguez-Garrote[24], Marcos Melián[25], Roberto Pazo-Cid[26], Mónica Guillot[27], David Gutiérrez[28], Ignacio García Escobar[29], Teresa Macarulla[6]

1. Medical Oncology Department. IGC Oncología Guipúzcoa. Spain
2. Medical Oncology Department. Hospital Universitario Central de Asturias, ISPA, Oviedo, Spain
3. Medical Oncology Department. 12 de Octubre University Hospital, Imas12, UCM, Madrid, Spain
4. Medical Oncology Department. Hospital General Universitario Gregorio Marañón, Universidad Complutense, Madrid, Spain
5. Department of Medical Oncology, Hospital Universitario de Salamanca, Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, Spain
6. Upper Gastrointestinal and Endocrine Tumor Unit, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital, Barcelona, Spain.
7. Department of Medical Oncology, IMIBIC, Cordoba University, CIBERONC. Reina Sofia University Hospital, Cordoba, Spain
8. Medical Oncology Department. Consorcio Hospital General Universitario, de Valencia, Spain
9. Department of Medical Oncology. Centro Integral Oncológico Clara Campal, HM Hospitales, Madrid, Spain
10. Department of Medical Oncology. Hospital Clínico San Carlos. Instituto de Investigación Hospital Clínico San Carlos (IdISSC), Universidad Complutense, Madrid, Spain
11. Medical Oncology Department. Complejo Hospitalario Universitario de Ourense
12. Medical Oncology Department. A Coruña University Hospital. Instituto Investigacion Biomedica INIBIC, A Coruña, Spain
13. Medical Oncology Department. Hospital Universitario Marqués de Valdecilla, IDIVAL, Santander. Spain
14. Department of Medical Oncology, Elche University Hospital, Alicante, Spain
15. Medical Oncology Department. Hospital Universitario de Navarra. Spain

16. Medical Oncology Department. UGCI Medical Oncology. University Regional and Virgen Victoria Hospital. IBIMA. Malaga. Spain
17. Department of Medical Oncology, Virgen del Rocio University Hospital, IBIS, Sevilla, Spain
18. Department of Medical Oncology, Hospital General University Santa Lucia, Cartagena, Murcia, Spain
19. Department of Medical Oncology, Hospital University La Paz, Madrid, Spain
20. Spanish Cooperative Group for the Treatment of Digestive Tumors (TTD). Madrid. Spain.
21. Catalan Institute of Oncology (ICO), L'Hospitalet de Llobregat, Barcelona, Spain.
22. Department of Medical Oncology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
23. Department of Medical Oncology, Hospital Universitario Príncipe de Asturias, Madrid, Spain.
24. Department of Medical Oncology, Hospital Ramón y Cajal, Madrid, Spain.
25. Department of Medical Oncology, Instituto Valenciano de Oncología, Valencia, Spain.
26. Department of Medical Oncology, Hospital Universitario Miguel Servet, Zaragoza, Spain.
27. Department of Medical Oncology, Hospital Universitario Son Espases, Mallorca, Spain.
28. Department of Medical Oncology, Hospital de Fuenlabrada, Madrid, Spain.
29. Department of Medical Oncology, Hospital Universitario de Toledo, Toledo, Spain.

DURVABTC Group

Margherita Rimini[1], Lorenzo Fornaro[2], Mario Domenico Rizzato[3], Lorenzo Antonuzzo[4-5], Federico Rossari[1], Tomoyuki Satake[6], Hanne Vandeputte[7], Caterina Vivaldi[2-8], Tiziana Pressiani[9], Jessica Lucchetti[10], Jin Won Kim[11], Oluseyi Abidoye[12], Ilario Giovanni Rapposelli[13], Stefano Tamberi[14], Fabian Finkelmeier[15], Guido Giordano[16-17], Federico Nichetti[18], Hong Jae Chon[19], Chiara Braconi[20], Chiara Pirrone[21], Florian Castet[22], Emiliano Tamburini[23], Changhoon Yoo[24], Alessandro Parisi[25], Anna Diana[26], Mario Scartozzi[27], Gerald W. Prager[28], Antonio Avallone[29], Marta Schirripa[30], Il Hwan Kim[31], Lukas Perkhofer[32-33], Ester Oneda[34], Monica Verrico[35], Jorge Adeva[36], Stephen L. Chan[37], Gian Paolo Spinelli[38], Nicola Personeni[39], Ingrid Garajova[40], Maria Grazia Rodriquenz[41], Silvana Leo[42], Francesca Salani[2-8], Antonio De Rosa[3,43], Daniele Lavacchi[4], Silvia Foti[1], Masafumi Ikeda[6], Jeroen Dekervel[7], Monica Niger[18], Rita Balsano[9-44], Giuseppe Tonini[45], Minsu Kang[11], Tanios Bekaii-Saab[12], Luca Esposito[13], Alessandra Boccaccino[14], Vera Himmelsbach[15], Matteo Landriscina[16-17], Selma Ahcene Djaballah[3], Valentina Zanuso[9-44], Gianluca Masi[2,8], Sara Lonardi[3], Lorenza Rimassa[9,44], Andrea Casadei-Gardini[1]

1 Department of Oncology, Vita-Salute San Raffaele University, IRCCS San Raffaele Scientific Institute Hospital, Milan, Italy
2 Unit of Medical Oncology 2, Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy
3 Department of Oncology, Veneto Institute of Oncology IOV - IRCCS, Padua, Italy
4 Clinical Oncology Unit, Careggi University Hospital, Florence, Italy; Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy.
5 Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy; Thoracic Surgery Unit, Careggi University Hospital, Florence, Italy.
6 Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital East, Kashiwa, Japan
7 Digestive Oncology, University Hospitals Leuven, Leuven, Belgium
8 Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
9 Medical Oncology and Hematology Unit, Humanitas Cancer Center, IRCCS Humanitas Research Hospital, Rozzano (Milan), Italy
10 Operative Research Unit of Medical Oncology, Fondazione Policlinico Universitario Campus Bio-Medico, Via Alvaro del Portillo, 200 - 00128 Roma, Italy
11 Division of Hematology/Medical Oncology, Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Gumi-ro 173 Beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 13620, Republic of Korea
12 Department of Internal Medicine, Mayo Clinic, Phoenix, AZ, USA
13 Department of Medical Oncology, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
14 Medical Oncology, Santa Maria delle Croci hospital, Ravenna AUSL Romagna ITALY
15 Medical Clinic 1, University Hospital, Goethe-University Frankfurt, Frankfurt am Main, Germany
16 Unit of Medical Oncology and Biomolecular Therapy, Policlinico Riuniti, Foggia
17 Department of Medical and Surgical Sciences, University of Foggia
18 Department of Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.
19 Division of Medical Oncology, Department of Internal Medicine, CHA Bundang Medical Center, CHA University School of Medicine, Seongnam, South Korea
20 University of Glasgow (School of Cancer Sciences), Beatson West of Scotland Cancer Centre, CRUK Scotland Centre
21 IRCCS Ospedale Policlinico San Martino, Medical Oncology Unit 1, Genova, Italy
22 Upper Gastrointestinal and Endocrine Tumor Unit, Vall d'Hebron Institute of Oncology (VHIO), Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain.
23 Oncology Department and Palliative Care, Cardinale Panico Tricase City Hospital, 73039 Tricase, Italy.
24 ASAN Medical Center, University of Ulsan College of Medicine
25 Clinica Oncologica e Centro Regionale di Genetica Oncologica, Università Politecnica delle Marche, Azienda Ospedaliero-Universitaria delle Marche, Via Conca 71, 60126 Ancona, Italy.
26 Oncology Unit, Ospedale del Mare, Napoli
27 Department of Biomedical Sciences, Unit of Biology and Genetics, University of Cagliari, Cagliari, Italy.
28 Department of Medicine I, Clinical Division of Oncology, Medical University Vienna, Austria
29 Clinical Experimental Abdominal Oncology Unit, Istituto Nazionale Tumori-IRCCS Fondazione G. Pascale, 80131, Naples, Italy.
30 Medical Oncology Unit, Department of Oncology and Hematology, Belcolle Hospital, Viterbo, Italy
31 Division of Oncology, Department of Internal Medicine, Haeundae Paik Hospital, Inje University College of Medicine, Busan, Republic of Korea.
32 Internal Medicine 1, University Hospital Ulm, Ulm, Germany
33 Institute of Molecular Oncology and Stem Cell Biology, Ulm University Hospital, Ulm, Germany
34 Dipartimento di Oncologia medica, Fondazione Poliambulanza, Brescia
35 UOC Oncologia A, Department of Hematology, Oncology and Dermatology, Policlinico Umberto I University Hospital, Sapienza University o f Rome, Viale Regina Elena, 324, 00161 Rome, Italy
36 12 de Octubre University Hospital, Spanish Society of Medical Oncology (SEOM), Madrid, Spain.
37 Department of Clinical Oncology, Prince of Wales Hospital, The Chinese University of Hong Kong
38 UOC Oncologia Territoriale, Polo Pontino, La Sapienza Università Di Roma, Latina, Italy.
39 Medical Oncology Unit, P.O. Manerbio - ASST Garda, 25025 Manerbio (Brescia)
40 Medical Oncology Unit, University Hospital of Parma, 43126 Parma, Italy
41 Oncology Unit, Fondazione IRCCS "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Italy
42 Division of Oncology, Vito Fazzi Hospital, Lecce, Italy.
43 Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy
44 Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (Milan), Italy
45 Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo, 21 - 00128 Roma, Italy

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2025.05.020.

References

Author names in bold designate shared co-first authorship

[1] Valle JW, Kelley RK, Nervi B, et al. Biliary tract cancer. Lancet 2021;397:428–444. https://doi.org/10.1016/S0140-6736(21)00153-7.
[2] Izquierdo-Sanchez L, Lamarca A, La Casta A, et al. Cholangiocarcinoma landscape in Europe: diagnostic, prognostic and therapeutic insights from the ENSCCA Registry. J Hepatol 2022;76:1109–1121. https://doi.org/10.1016/j.jhep.2021.12.010.

[3] Oh D-Y, Ruth He A, Qin S, et al. Durvalumab plus gemcitabine and cisplatin in advanced biliary tract cancer. NEJM Evid 2022;1. https://doi.org/10.1056/EVIDoa2200015.

[4] Kelley RK, Ueno M, Yoo C, et al. Pembrolizumab in combination with gemcitabine and cisplatin compared with gemcitabine and cisplatin alone for patients with advanced biliary tract cancer (KEYNOTE-966): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2023;401:1853–1865. https://doi.org/10.1016/S0140-6736(23)00727-4.

[5] Merino M, Kasamon Y, Theoret M, et al. Irreconcilable differences: the divorce between response rates, progression-free survival, and overall survival. J Clin Oncol 2023;41:2706–2712. https://doi.org/10.1200/JCO.23.00225.

[6] Saad ED, Buyse M. Statistical controversies in clinical research: end points other than overall survival are vital for regulatory approval of anticancer agents. Ann Oncol 2016;27:373–378. https://doi.org/10.1093/annonc/mdv562.

[7] Korn EL, Freidlin B, Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. J Clin Oncol 2011;29:2439–2442. https://doi.org/10.1200/JCO.2011.34.6056.

[8] Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. Int J Technol Assess Health Care 2014;30:312–324. https://doi.org/10.1017/S0266462314000300.

[9] Elbaz J, Haslam A, Prasad V. An empirical analysis of overall survival in drug approvals by the US FDA (2006–2023). Cancer Med 2024;13. https://doi.org/10.1002/cam4.7190.

[10] Naci H, Zhang Y, Woloshin S, et al. Overall survival benefits of cancer drugs initially approved by the US Food and Drug Administration on the basis of immature survival data: a retrospective analysis. Lancet Oncol 2024;25:760–769. https://doi.org/10.1016/S1470-2045(24)00152-9.

[11] Neuzillet C, Malka D, Lièvre A, et al. Correlation between efficacy endpoints in patients with advanced biliary tract cancer treated by systemic second-line therapies: analysis of aggregated data from a systematic literature review. Clin Res Hepatol Gastroenterol 2022;46:102010. https://doi.org/10.1016/j.clinre.2022.102010.

[12] Moriwaki T, Yamamoto Y, Gosho M, et al. Correlations of survival with progression-free survival, response rate, and disease control rate in advanced biliary tract cancer: a meta-analysis of randomised trials of first-line chemotherapy. Br J Cancer 2016;114:881–888. https://doi.org/10.1038/bjc.2016.83.

[13] Lamarca A, Hubner RA, David Ryder W, et al. Second-line chemotherapy in advanced biliary cancer: a systematic review. Ann Oncol 2014;25:2328–2338. https://doi.org/10.1093/annonc/mdu162.

[14] Eckel F, Schmid RM. Chemotherapy in advanced biliary tract carcinoma: a pooled analysis of clinical trials. Br J Cancer 2007;96:896–902. https://doi.org/10.1038/sj.bjc.6603648.

[15] Belin L, Tan A, De Rycke Y, et al. Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. Br J Cancer 2020;122:1707–1714. https://doi.org/10.1038/s41416-020-0805-y.

[16] Lamarca A, Macarulla T. Facts and hopes in the systemic therapy of biliary tract carcinomas. Clin Cancer Res 2024. https://doi.org/10.1158/1078-0432.CCR-22-2438.

[17] Xie W, Halabi S, Tierney JF, et al. A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses. JNCI Cancer Spectr 2019;3. https://doi.org/10.1093/jncics/pkz002.

[18] Gharzai LA, Jiang R, Wallington D, et al. Intermediate clinical endpoints for surrogacy in localised prostate cancer: an aggregate meta-analysis. Lancet Oncol 2021;22:402–410. https://doi.org/10.1016/S1470-2045(20)30730-0.

[19] Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Med 2017;15:134. https://doi.org/10.1186/s12916-017-0902-9.

[20] Valle JW, Wasan H, Johnson P, et al. Gemcitabine alone or in combination with cisplatin in patients with advanced or metastatic cholangiocarcinomas or other biliary tract tumours: a multicentre randomised phase II study – the UK ABC-01 Study. Br J Cancer 2009;101:621–627. https://doi.org/10.1038/sj.bjc.6605211.

[21] Valle J, Wasan H, Palmer DH, et al. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. N Engl J Med 2010;362:1273–1281. https://doi.org/10.1056/NEJMoa0908721.

[22] Valle JW, Wasan H, Lopes A, et al. Cediranib or placebo in combination with cisplatin and gemcitabine chemotherapy for patients with advanced biliary tract cancer (ABC-03): a randomised phase 2 trial. Lancet Oncol 2015;16:967–978. https://doi.org/10.1016/S1470-2045(15)00139-4.

[23] Yoo C, Kim K, Jeong JH, et al. Liposomal irinotecan plus fluorouracil and leucovorin versus fluorouracil and leucovorin for metastatic biliary tract cancer after progression on gemcitabine plus cisplatin (NIFTY): a multicentre, open-label, randomised, phase 2b study. Lancet Oncol 2021;22:1560–1572. https://doi.org/10.1016/S1470-2045(21)00486-1.

[24] **Choi IS**, **Kim KH**, Lee JH, et al. A randomised phase II study of oxaliplatin/5-FU (mFOLFOX) versus irinotecan/5-FU (mFOLFIRI) chemotherapy in locally advanced or metastatic biliary tract cancer refractory to first-line gemcitabine/cisplatin chemotherapy. Eur J Cancer 2021;154:288–295. https://doi.org/10.1016/j.ejca.2021.06.019.

[25] Ramsey SD, Onar-Thomas A, Wheeler SB. Real-world database studies in oncology: a call for standards. J Clin Oncol 2024;42:977–980. https://doi.org/10.1200/JCO.23.02399.

[26] Booth CM, Eisenhauer EA, Gyawali B, et al. Progression-free survival should not Be used as a primary end point for registration of anticancer drugs. J Clin Oncol 2023. https://doi.org/10.1200/JCO.23.01423.

[27] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021:n71. https://doi.org/10.1136/bmj.n71.

[28] Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for systematic reviews of interventions version 6.4. Cochrane; 2023 (updated August 2023).

[29] Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi List: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol 1998;51:1235–1241. https://doi.org/10.1016/S0895-4356(98)00131-0.

[30] Johnson KR, Ringland C, Stokes BJ, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncol 2006;7:741–746. https://doi.org/10.1016/S1470-2045(06)70800-2.

[31] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5:173–186. https://doi.org/10.1002/pst.207.

[32] Conforti F, Pala L, Sala I, et al. Evaluation of pathological complete response as surrogate endpoint in neoadjuvant randomised clinical trials of early stage breast cancer: systematic review and meta-analysis. BMJ 2021:e066381. https://doi.org/10.1136/bmj-2021-066381.

[33] Schemper M, Kaider A, Wakounig S, et al. Estimating the correlation of bivariate failure times under censoring. Stat Med 2013;32:4781–4790. https://doi.org/10.1002/sim.5874.

[34] Eden SK, Li C, Shepherd BE. Nonparametric estimation of Spearman's rank correlation with bivariate survival data. Biometrics 2022;78:421–434. https://doi.org/10.1111/biom.13453.

[35] Norsworthy KJ, Gao X, Ko C-W, et al. Response rate, event-free survival, and overall survival in newly diagnosed acute myeloid leukemia: US food and drug administration trial-level and patient-level analyses. J Clin Oncol 2022;40:847–854. https://doi.org/10.1200/JCO.21.01548.

[36] Blumenthal GM, Karuri SW, Zhang H, et al. Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non–small-cell lung cancer: US food and drug administration trial-level and patient-level analyses. J Clin Oncol 2015;33:1008–1014. https://doi.org/10.1200/JCO.2014.59.0489.

[37] Simon R, Makuch RW. A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. Stat Med 1984;3:35–44. https://doi.org/10.1002/sim.4780030106.

[38] Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. J Clin Oncol 1983;1:710–719. https://doi.org/10.1200/JCO.1983.1.11.710.

[39] Prasad V, Kim C, Burotto M, et al. The strength of association between surrogate end points and survival in oncology. JAMA Intern Med 2015;175:1389. https://doi.org/10.1001/jamainternmed.2015.2829.

[40] Abou-Alfa GK, Macarulla T, Javle MM, et al. Ivosidenib in IDH1-mutant, chemotherapy-refractory cholangiocarcinoma (ClarIDHy): a multicentre, randomised, double-blind, placebo-controlled, phase 3 study. Lancet Oncol 2020;21:796–807. https://doi.org/10.1016/S1470-2045(20)30157-1.

[41] Knox J, Bazin I, Oh D, et al. O-2 Phase III study of NUC-1031 + cisplatin vs gemcitabine + cisplatin for first-line treatment of patients with advanced biliary tract cancer (NuTide:121). Ann Oncol 2023;34:S180–S181. https://doi.org/10.1016/j.annonc.2023.04.017.

[42] Lassere MN, Johnson KR, Schiff M, et al. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? an analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-

variables regression, the surrogate threshold effect (STE) and the biomarker-surrogacy (BioSurrogate) evaluation schema (BSES). BMC Med Res Methodol 2012;12:27. https://doi.org/10.1186/1471-2288-12-27.
[43] Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res 2008;17:303–340. https://doi.org/10.1177/0962280207082719.
[44] Institute for Quality and Efficiency in Health Care (IQWiG). Validity of surrogate endpoints in oncology: executive summary of rapid report A10-05, Version 1.1. Inst Qual Efficiency Health Care Executive Summ 2011.
[45] Garnick MB. Preserving the sanctity of overall survival for drugs approved on the basis of progression-free survival: tivozanib as a case study. J Clin Oncol 2013;31:3746–3748. https://doi.org/10.1200/JCO.2013.51.4869.
[46] Cherny NI, Dafni U, Bogaerts J, et al. ESMO-magnitude of clinical benefit Scale version 1.1. Ann Oncol 2017;28:2340–2366. https://doi.org/10.1093/annonc/mdx310.
[47] Tannock IF, Pond GR, Booth CM. Biased evaluation in cancer drug trials—how use of progression-free survival as the primary end point can mislead. JAMA Oncol 2022;8:679. https://doi.org/10.1001/jamaoncol.2021.8206.
[48] Fojo T, Simon RM. Inappropriate censoring in Kaplan-Meier analyses. Lancet Oncol 2021;22:1358–1360. https://doi.org/10.1016/S1470-2045(21)00473-3.
[49] Gilboa S, Pras Y, Mataraso A, et al. Informative censoring of surrogate endpoint data in phase 3 oncology trials. Eur J Cancer 2021;153:190–202. https://doi.org/10.1016/j.ejca.2021.04.044.
[50] Walia A, Tuia J, Prasad V. Progression-free survival, disease-free survival and other composite end points in oncology: improved reporting is needed. Nat Rev Clin Oncol 2023;20:885–895. https://doi.org/10.1038/s41571-023-00823-5.
[51] Harry VN, Semple SI, Parkin DE, et al. Use of new imaging techniques to predict tumour response to therapy. Lancet Oncol 2010;11:92–102. https://doi.org/10.1016/S1470-2045(09)70190-1.
[52] Sahani DV, Hayano K, Galluzzo A, et al. Measuring treatment response to systemic therapy and predicting outcome in biliary tract cancer: comparing tumor size, volume, density, and metabolism. AJR Am J Roentgenol 2015;204:776–781. https://doi.org/10.2214/AJR.14.13223.
[53] Inker LA, Collier W, Greene T, et al. A meta-analysis of GFR slope as a surrogate endpoint for kidney failure. Nat Med 2023;29:1867–1876. https://doi.org/10.1038/s41591-023-02418-0.
[54] Valle JW, Vogel A, Denlinger CS, et al. Addition of ramucirumab or merestinib to standard first-line chemotherapy for locally advanced or metastatic biliary tract cancer: a randomised, double-blind, multicentre, phase 2 study. Lancet Oncol 2021;22:1468–1482. https://doi.org/10.1016/S1470-2045(21)00409-5.
[55] Kataria B, Sharma A, Pramanik R, et al. PD-9 Three-arm phase II/III randomized controlled trial in patients with unresectable/metastatic gall bladder cancer with poor performance status: Erlotinib or capecitabine v/s best supportive care. Ann Oncol 2022;33:S242. https://doi.org/10.1016/j.annonc.2022.04.087.
[56] Sharma A, Dwary AD, Mohanti BK, et al. Best supportive care compared with chemotherapy for unresectable gall bladder cancer: a randomized controlled study. J Clin Oncol 2010;28:4581–4586. https://doi.org/10.1200/JCO.2010.29.3605.
[57] Assenat E, Blanc JF, Bouattour M, et al. 48P (BREGO) Regorafenib combined with modified m-GEMOX in patients with advanced biliary tract cancer (BTC): a phase II randomized trial. Ann Oncol 2021;32:S376–S377. https://doi.org/10.1016/j.annonc.2021.08.327.

Keywords: Cholangiocarcinoma; Biliary tract cancer; Surrogate endpoints; Trial design; Overall survival; Meta-analysis.
Received 9 December 2024; received in revised form 18 April 2025; accepted 6 May 2025; available online 3 June 2025