Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Antiangiogenic Class and Trial Population Are Confounded in First- Line Immunotherapy for Hepatocellular Carcinoma: A Bayesian Network Meta-Analysis of Nine Randomised Trials

Burkhoniddin Tillashaykhov
Central Asian University
Iroda Ikromova
Central Asian University
Dilnavoz Berdiyeva
d.berdiyeva@centralasian.uz

Central Asian University

Research Article

Keywords:

Posted Date: October 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11367042/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Gastroenterology on October 31st, 2026. See the published version at https://doi.org/10.1186/s12876-026-05496-9.

Research

Antiangiogenic Class and Trial Population Are Confounded in First-Line Immunotherapy for Hepatocellular Carcinoma: A Bayesian Network Meta-Analysis of Nine Randomised Trials

Burkhoniddin Tillashaykhov, Iroda Ikromova, Dilnavoz Berdiyeva*

Central Asian University, Tashkent, Uzbekistan

*** Corresponding author.** Dilnavoz Berdiyeva, PhD, Associate Professor, Central Asian University, School of Medicine, 264 Milliy Bog St, Tashkent 111221, Uzbekistan. Tel.: +998 90 926 0509. E-mail: d.berdiyeva@centralasian.uz

ORCID. Burkhoniddin Tillashaykhov 0009-0003-4865-0619; Iroda Ikromova 0009-0001-3096-6397; Dilnavoz Berdiyeva 0009-0002-9668-0479

Article type: Research (systematic review and network meta-analysis) | Main text: ~5,550 words | Tables: 2 | Figures: 3 | References: 31

PROSPERO registration: CRD420261353528

Abstract

Background First-line therapy for advanced hepatocellular carcinoma (HCC) combines an immune checkpoint inhibitor (ICI) with one of two antiangiogenic backbones: an anti-VEGF monoclonal antibody or a VEGFR-targeted tyrosine kinase inhibitor (TKI). No head-to-head trial exists, and successive network meta-analyses have ranked the two classes and reported class-level probabilities as estimates of a drug-class property.

Objective To estimate the class comparison and, separately, to test whether a class effect is identifiable from the aggregate trial data on which such rankings are based.

Methods Nine randomised trials (5,333 patients) were synthesised in a Bayesian random-effects network anchored on sorafenib. Because each treatment appears in a single trial, between-study heterogeneity is not identifiable from the data; it was therefore specified using the external empirical predictive distributions of Turner et al., and every efficacy estimate reported here is conditional on that externally sourced prior. Outcomes were overall survival (OS), progression-

free survival (PFS), objective response (ORR) and grade ≥3 adverse events. Five effect modifiers were pre-specified; their joint distribution across trials was examined before interpreting any class-level estimate.

Results Three regimens improved OS over sorafenib: sintilimab plus bevacizumab (HR 0.57, 95% credible interval 0.35–0.92), finotonlimab plus bevacizumab (0.60, 0.37–0.99) and camrelizumab plus rivoceranib (0.64, 0.41–0.99). The posterior probability favouring the anti-VEGF-antibody class was 75.0% for OS, 74.8% for ORR and 45.3% for PFS. These estimates are not interpretable as class effects. Treatment class, hepatitis B virus prevalence, macrovascular invasion and China-only enrolment were mutually collinear across trials (pairwise r = 0.81–0.91; HBV–China r = 0.87), so a survival advantage attributed to drug class is equally consistent with one attributable to the populations in which each class was tested. With nine trials, trial-level meta-regression is underpowered to separate the independent contributions of viral aetiology and geographic enrolment and is reported as hypothesis-generating only; no reweighting or prior specification separates them either, and sensitivity analyses moved the class probability only between 73.9% and 85.5% without altering this. The contrasts that do survive are those estimated within randomised comparisons: grade ≥3 hand-foot syndrome occurred with every TKI regimen (2–12%) and with none of the four anti-VEGF-antibody regimens, against 6.9–12% in the sorafenib arms of those same trials.

Conclusions Class rankings in first-line HCC describe the evidence base rather than a property of the drugs. The collinearity of treatment class with viral aetiology and geographic region makes class attribution unidentifiable from aggregate data, irrespective of model choice, and this limitation applies to the published rankings as much as to ours. On GRADE-NMA the class-level indirect comparison was rated very low certainty, downgraded for unresolvable confounding by trial population. Separating drug class from trial population will require individual-patient-data analysis or head-to-head randomisation; until then, within-trial contrasts remain the defensible basis for class-level claims.

Keywords Hepatocellular carcinoma; Network meta-analysis; Immune checkpoint inhibitors; Bevacizumab; Tyrosine kinase inhibitors; Confounding; Identifiability; Bayesian analysis

1 Introduction

Liver cancer is the sixth most frequently diagnosed cancer worldwide and the third leading cause of cancer death, with an estimated 865,000 new cases and 758,000 deaths in 2022 [1]; hepatocellular carcinoma (HCC) accounts for the large majority of cases, and most patients present at a stage where curative treatment is no longer possible. A further proportion reach advanced disease through recurrence after curative-intent resection, where preoperative imaging can flag pathological features associated with relapse, such as microvascular invasion and satellite nodules [2], and where variation in operative strategy affects perioperative outcome without altering long-term survival [3].

First-line systemic therapy consisted for more than a decade of sorafenib [4] and later lenvatinib [5]. Since IMbrave150 [6,7], the standard has been an immune checkpoint inhibitor (ICI) combined with an antiangiogenic agent, and a succession of phase 3 trials has extended the approach across a growing range of regimens [8–15], an approach now embedded in national treatment guidelines [16]. These combinations pair a checkpoint inhibitor with one of two mechanistically distinct partners: a monoclonal antibody targeting vascular endothelial growth factor (VEGF), as in atezolizumab plus bevacizumab [6], sintilimab plus a bevacizumab biosimilar [8], toripalimab plus bevacizumab [12] or finotonlimab plus bevacizumab [13]; or an orally administered multikinase inhibitor of the VEGF receptor (VEGFR), as in camrelizumab plus rivoceranib [15], atezolizumab plus cabozantinib [9,11], pembrolizumab plus lenvatinib [10] or anlotinib plus penpulimab [14].

No trial has compared the two backbones directly. In the absence of head-to-head randomisation, three network meta-analyses have compared them indirectly — across six trials [17], stratified by viral aetiology [18], and at strategy level across fifteen trials [19] — and all report rankings favouring the anti-VEGF-antibody class. These rankings now circulate as evidence about the drugs: as a statement that antibody-based blockade of VEGF is the more effective way to partner a checkpoint inhibitor in this disease.

That interpretation rests on a specific assumption. An indirect class comparison estimates a drug-class effect only if the trials contributing to each class are exchangeable with respect to the factors that modify treatment benefit. This requirement, and the conditions under which indirect comparison becomes uninterpretable, are well established in the network meta-analysis

methodology literature: the validity of an indirect estimate depends on how effect modifiers are distributed across the contributing trials [20], and transitivity and identifiability are recognised as conceptual rather than computational constraints [21,22]. In first-line HCC this condition is strained in a specific and checkable way: the trials are concentrated by geography and by viral aetiology, and that concentration is not independent of which backbone each regimen used. If treatment class and population characteristics vary together across the evidence base, then a class ranking and a population ranking are numerically the same object, and no amount of modelling will tell them apart.

We therefore performed a Bayesian network meta-analysis of first-line randomised trials in advanced HCC with two aims. The first was to estimate the comparative efficacy and toxicity of the two antiangiogenic classes. The second, which we treat as the primary contribution, was to establish whether a class effect is identifiable at all from aggregate data of this structure, by examining the joint distribution of pre-specified effect modifiers across trials and by testing how far the class comparison can be moved by model specification. We report both, and we set out which claims the evidence base does and does not support. We note at the outset that the design which would settle the question is unlikely to be run. A head-to-head trial powered to detect a hazard ratio of approximately 0.89, the magnitude separating the best-performing regimen in each class in this analysis, would require some 2,300 deaths and therefore, at the overall-survival maturity reached in the included trials, between roughly 1,500 and 1,800 patients per arm; and no sponsor has a commercial incentive to risk its own regimen against a competitor's. The feasible alternative, pooled individual-patient-data analysis, would require every sponsor of the nine trials to release patient-level covariates and outcomes to a common analysis, which would permit the aetiology-by-treatment interaction to be estimated within trials rather than across them.

2 Materials and Methods

2.1 Protocol, Registration, and Reporting

This systematic review and Bayesian network meta-analysis (NMA) was conducted and reported in accordance with the PRISMA 2020 statement [23] and the PRISMA extension for network meta-analyses [24]. The protocol was prospectively registered with PROSPERO (CRD420261353528). The study selection process is shown in Supplementary Figure S1.

2.2 Eligibility Criteria

We included phase 2/3 randomised controlled trials of first-line systemic therapy in adults with advanced or unresectable HCC in which an ICI was combined with an antiangiogenic agent (either an anti-VEGF monoclonal antibody or a VEGFR-targeted TKI) and compared with an active systemic comparator, and which reported at least one of the outcomes of interest. Trials of dual checkpoint blockade without an antiangiogenic partner, trials in the adjuvant, neoadjuvant, or later-line settings, trials combining systemic therapy with locoregional treatment, single-arm studies, and post-hoc or subgroup reports without independent randomised data were excluded.

2.3 Information Sources and Search

We searched PubMed/MEDLINE and the Cochrane Central Register of Controlled Trials (CENTRAL) and screened the ClinicalTrials.gov registry for completed trials; no additional eligible records were identified beyond those retrieved from the two databases. Data extraction and analysis followed completion of study selection. Although the protocol listed additional databases, all eligible phase 3 trials were captured by PubMed and CENTRAL, and no further eligible records were identified from other sources. Full search strategies are given in the Supplementary Material.

2.4 Study Selection and Data Extraction

Records were screened independently by two reviewers against the eligibility criteria, with disagreements resolved by discussion. For each included trial we extracted, in duplicate, the trial identifier and registration number, design and phase, arm-level sample sizes, regimens and comparators, median follow-up, and the primary efficacy estimates (OS hazard ratio [HR] and 95% confidence interval [CI], PFS HR and 95% CI, and objective response rate [ORR] in each arm). We additionally extracted the pre-specified effect-modifier variables (see below) and prognostic characteristics (Eastern Cooperative Oncology Group performance status, Child–Pugh class, macrovascular invasion or extrahepatic spread [MVI/EHS], and alpha-fetoprotein [AFP] ≥400 ng/mL). Every extracted value was verified against the primary trial publication, and discrepancies identified during verification were corrected to the source value. For trials reported at more than one analysis stage, the most mature analysis was used: the COSMIC-312 final overall-survival analysis [11] replaced the earlier interim assessment, and the CARES-310 final analysis [15] replaced the 2023 interim report.

2.5 Effect Modifiers (Pre-Specified)

Five candidate effect modifiers were specified before examining any outcome data: hepatitis B virus (HBV) prevalence, AFP ≥400 ng/mL, MVI/EHS, ECOG performance status, and Child–Pugh class.

HBV prevalence was selected as the primary modifier on biological and practical grounds: hepatitis B-related and non-viral HCC differ in their response to immune checkpoint inhibition, and HBV status was reported consistently across all nine trials. The proportion of Asian patients was considered but defined inconsistently across trials, some by geographic region and others by race, and was therefore treated as a descriptive characteristic only. HBV prevalence is not independent of treatment class in this network: the four trials with the highest HBV prevalence were all conducted in China and contribute a specific mix of regimens, so any apparent HBV-related modification of effect is confounded with class composition and cannot be interpreted causally. The associated meta-regression is therefore exploratory.

2.6 Network Geometry and the Lenvatinib Bridge

The evidence network was anchored on sorafenib, the common comparator shared by the majority of trials. One trial (LEAP-002) used lenvatinib rather than sorafenib as its control; lenvatinib was connected to the sorafenib anchor through the REFLECT trial (lenvatinib versus sorafenib [5]), forming a single two-link indirect bridge. The pembrolizumab plus lenvatinib node is therefore connected to the rest of the network only through this bridge, and is the one regimen in the network whose estimate is two indirect steps from the common anchor rather than one. To quantify the cost of that extra step we refitted the overall-survival model with the LEAP-002 contrast re-anchored directly on sorafenib, holding its magnitude and standard error fixed, and compared the posterior precision of the pembrolizumab plus lenvatinib estimate between the two fits. Because the resulting network is a star with a single appended bridge and contains no closed loops, statistical evaluation of consistency (node-splitting or the Bucher method) is structurally inapplicable. Transitivity was instead assessed by comparing the distribution of the pre-specified effect modifiers across trials (Table 1).

2.7 Statistical Analysis

Analyses were performed within a Bayesian framework. For the time-to-event outcomes (OS and PFS), trial-level log-hazard ratios and their standard errors were derived from the published hazard ratios and 95% confidence intervals and modelled with a normal likelihood. For ORR and

for grade ≥3 treatment-related adverse events, arm-level event counts were reconstructed from the reported proportions and arm sizes and modelled on the log-odds scale, with a continuity correction of 0.5 applied to all cells. Treatment discontinuation due to adverse events was pre-specified in the protocol but not analysed, as reporting definitions varied across trials and consistent extraction was not possible.

All endpoints were analysed with random-effects network meta-analysis models using the contrast-based parameterisation [25] on the appropriate scale, with the lenvatinib-comparator trial handled through the REFLECT bridge as described above so that every regimen was expressed relative to the common sorafenib reference.

Vague priors were placed on the basic parameters (normal, mean 0, standard deviation 10). Between-study heterogeneity was specified using empirical predictive distributions (see below). Posterior distributions were estimated by Hamiltonian Monte Carlo using the No-U-Turn Sampler in PyMC v5.28.5 [26], with four chains of 8,000 draws each after 6,000 tuning iterations, a fixed random seed, a target acceptance probability of 0.999, and a maximum tree depth of 15. Convergence was assessed by the Gelman–Rubin statistic (R-hat), effective sample size, and the absence of divergent transitions; all reported models achieved R-hat < 1.01, effective sample sizes above 10,000, and no divergent transitions.

Treatments were compared through their posterior hazard ratios (or odds ratios) versus sorafenib and through all pairwise contrasts (league table, Supplementary Table S5). Treatment rankings were summarised using the surface under the cumulative ranking curve (SUCRA) [27], interpreted alongside the full credible intervals. SUCRA values are reported throughout as descriptions of the available evidence base, not as estimates of comparative drug or drug-class efficacy; in a star network a SUCRA value is a transformation of a single trial's point estimate against the shared comparator and inherits every limitation of that estimate.

2.8 Specification of Between-Study Heterogeneity

Between-study heterogeneity was specified using the empirical predictive distributions of Turner et al. [28], selected by outcome type (pharmacological versus pharmacological). These distributions were derived as log-normal priors for the between-study variance of binary outcomes on the log odds-ratio scale; we applied them to log hazard ratios as an approximation and examined sensitivity to this choice using alternative priors (Section 2.9). The full

specification, and its rationale in a star network where heterogeneity is not identifiable from the data, is given in the Supplementary Methods.

2.9 Sensitivity and Robustness Analyses

The robustness of the primary OS comparison was examined by a fixed-effect model, substitution of alternative empirical heterogeneity priors [28] (pharmacological versus placebo; cause-specific mortality/composite), restriction to high-HBV trials (≥80% and ≥70% prevalence), leave-one-trial-out re-estimation with removal of any orphaned trial, and removal of the REFLECT bridge with restriction to the seven trials having a direct sorafenib comparator (Supplementary Table S3). An exploratory random-effects meta-regression of the trial-level OS log-hazard ratio on HBV prevalence was performed across the sorafenib-referenced trials. This was interpreted for direction only, given the small number of trials and the mutual confounding of treatment class with HBV prevalence and geographic region.

2.10 Risk of Bias

Risk of bias was assessed independently by two reviewers using the Cochrane Risk of Bias 2 tool (RoB 2) [29] across its five domains: the randomisation process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Reviewers agreed on all judgements. No trial was judged at high risk of bias in any domain. Domain-level and overall judgements for each trial are presented in Supplementary Table S6.

2.11 Certainty of Evidence

Certainty of the OS network estimates was assessed using GRADE for network meta-analysis [30] across risk of bias, indirectness, imprecision, and publication bias; incoherence could not be evaluated statistically in a star network and was addressed through the transitivity assessment. Each comparison against sorafenib was rated and agreed by all authors (Supplementary Table S4); comparisons between active regimens, being wholly indirect, were low or very low certainty throughout. Because the class-level contrast between the anti-VEGF-antibody and VEGFR-TKI backbones is the principal comparison of this analysis, it was additionally rated as a comparison in its own right, for each of the three efficacy endpoints and for grade ≥3 adverse events (Supplementary Table S8).

2.12 Use of Generative Artificial Intelligence

Generative artificial intelligence was not used to design the study, extract data, generate or analyse data, or produce figures. All statistical analyses were performed by the authors in PyMC, and all extracted values were verified by the authors against the primary trial publications. A large language model (Claude, Anthropic) was used to assist with language editing of the manuscript text; the authors reviewed and edited all output and take full responsibility for the content.

3 Results

3.1 Study Selection and Network Geometry

The systematic search identified 665 records (275 from PubMed/MEDLINE and 390 from Cochrane CENTRAL). After removal of 233 duplicates, 432 records were screened by title and abstract, of which 421 were excluded. Eleven full-text articles were assessed for eligibility; two were excluded (one second-line/post-hoc analysis and one transarterial chemoembolisation-combination trial), leaving nine randomised controlled trials for inclusion in the network meta-analysis (Supplementary Figure S1, PRISMA-NMA flow diagram). The nine trials randomised a total of 5,333 patients: 2,640 to a first-line ICI plus antiangiogenic regimen, 1,739 to the sorafenib or lenvatinib control arm of those eight trials, and 954 to the bridging trial (REFLECT). Eight trials used sorafenib as the comparator and one (LEAP-002) used lenvatinib. The evidence network formed a star centred on sorafenib; lenvatinib was connected to the anchor only through the REFLECT trial, forming a single two-link bridge through which the pembrolizumab plus lenvatinib node joined the remainder of the network (Figure 1). Because the network contained no closed loops, statistical evaluation of inconsistency was not applicable, and transitivity was assessed through the distribution of effect modifiers across trials (Table 1).

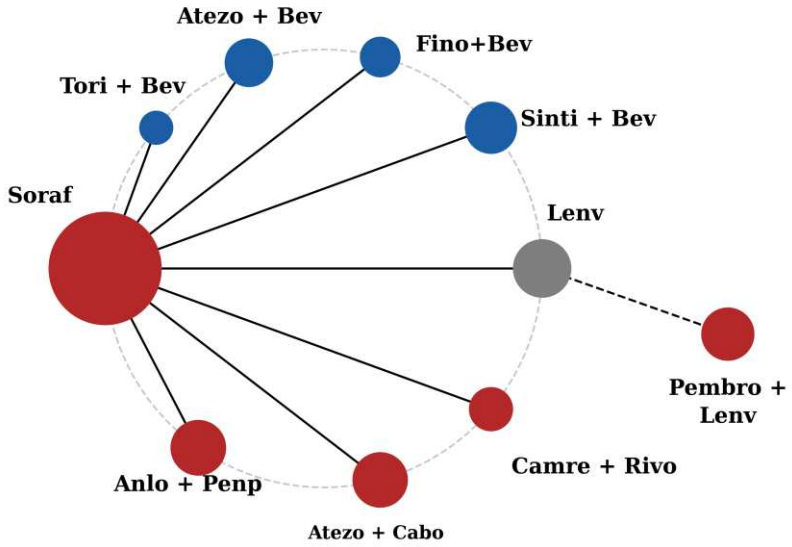


Figure 1 Network geometry of the evidence base. Nodes are treatment regimens; node area is proportional to the number of randomised participants. Edges denote direct comparisons. Node colour indicates class: blue, ICI plus anti-VEGF antibody; crimson, ICI plus TKI. Sorafenib (central) is the common comparator. Lenvatinib (grey) is an indirect bridge: pembrolizumab plus lenvatinib links to the network only via REFLECT and LEAP-002 (dashed edge). The star-shaped network has no closed loops, precluding assessment of direct–indirect consistency.

3.2 Trial and Patient Characteristics

The included trials were IMbrave150, CARES-310, ORIENT-32, COSMIC-312, LEAP-002, HEPATORCH, the finotonlimab trial, REFLECT, and APOLLO (Supplementary Table S1). The distribution of pre-specified effect modifiers is summarised in Table 1.

Table 1 Effect-modifier distribution across trials (transitivity assessment).

Trial	Class	HBV (%)	Non-Viral (%)	CP-A (%)	MVI/EHS (%)	AFP ≥400 (%)
ORIENT-32	anti-VEGF	94.2	3.3	95.8	79.3	43.1
Finotonlimab + bevacizumab	anti-VEGF	89.0	6.1	93.0	79.5	47.4
HEPATORCH	anti-VEGF	89.9	5.5	100	79.0	47.0
APOLLO	ICI + TKI	84.1	12.3	92.4	80.3	49.2
CARES-310	ICI + TKI	74.6	16.0	100	74.0	36.0
REFLECT	monotherapy	50.2	27.1	99.0	70.0	—

(bridge)						
LEAP-002	ICI + TKI	49.0	38.2	99.0	67.0	31.6
IMbrave150	anti-VEGF	47.9	30.5	99.0	75.0	37.0
COSMIC-312	ICI + TKI	29.4	39.0	100	69.0	35.1

Trials ordered by descending HBV prevalence. HBV prevalence (29–94%), the principal candidate effect modifier, was patterned by geography rather than treatment class: the four China-only trials had the highest values (84–94%), whereas globally enrolled trials, including the anti-VEGF trial IMbrave150 (47.9%), had the lowest. HBV is therefore partially confounded with both region and treatment class and cannot be read as a class-independent effect (see Section 4). Child–Pugh A (92–100%) and MVI/EHS (67–80%) were reasonably balanced; AFP ≥400 ng/mL varied more widely (32–49%). — REFLECT reported AFP at ≥200 ng/mL, not ≥400; the value is omitted as non-comparable. CP-A, Child–Pugh class A.

3.3 Overall Survival

Three regimens showed a credible survival advantage over sorafenib: sintilimab plus bevacizumab (HR 0.57, 95% credible interval [CrI] 0.35–0.92), finotonlimab plus bevacizumab (0.60, 0.37–0.99), and camrelizumab plus rivoceranib (0.64, 0.41–0.99); atezolizumab plus bevacizumab (0.66) and anlotinib plus penpulimab (0.69) had intervals marginally including unity, and atezolizumab plus cabozantinib showed no benefit (0.98, 0.62–1.54; COSMIC-312 final analysis [11]) (Figure 2). Heterogeneity was low (τ = 0.12). The posterior probability that the best anti-VEGF-antibody regimen was superior to the best TKI regimen was 75.0%.

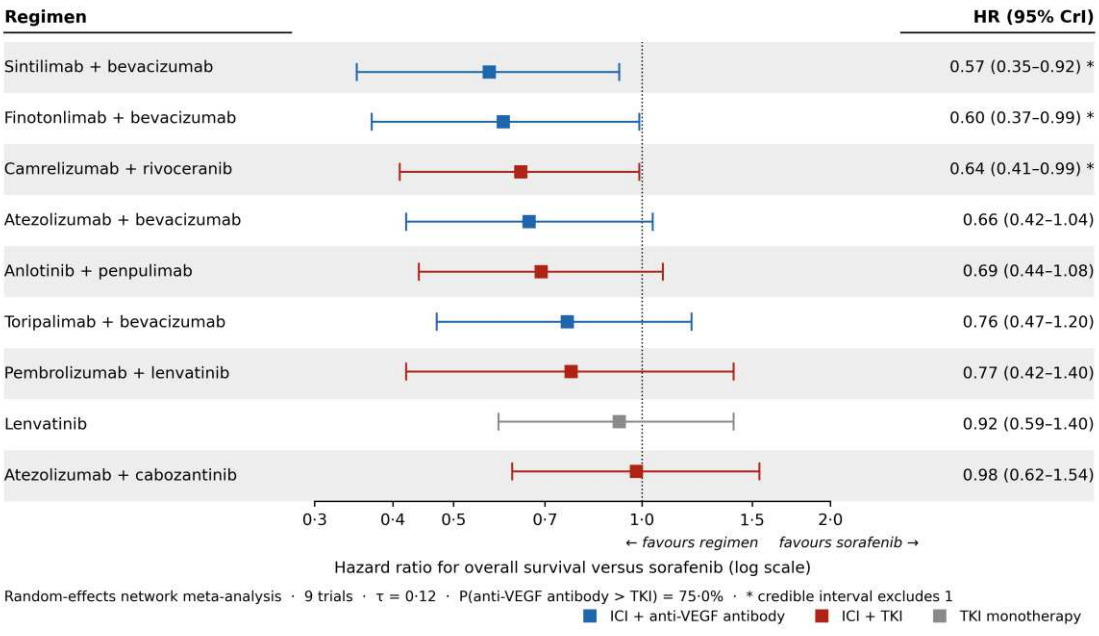


Figure 2 Overall survival: hazard ratios versus sorafenib. Posterior median hazard ratios (squares) with 95% credible intervals (horizontal lines) from the random-effects network meta-analysis, ordered from most to least

favourable. Colour denotes treatment class. Hazard ratios below 1 favour the experimental regimen over sorafenib. Asterisks mark the three regimens whose credible interval excluded unity. Random-effects model, nine trials, τ = 0.12; the posterior probability that the best anti-VEGF-antibody regimen was superior to the best TKI regimen was 75.0%.

3.4 Progression-Free Survival

For progression-free survival, the most favourable point estimates were finotonlimab plus bevacizumab (HR 0.50, 0.28–0.89), anlotinib plus penpulimab (0.52), and camrelizumab plus rivoceranib (0.54); three credible intervals excluded unity (τ = 0.14, Supplementary Table S2). Several TKI-based regimens ranked among the most effective, and the posterior probability favouring the anti-VEGF-antibody class was 45.3%, showing no class-level difference for this endpoint.

3.5 Objective Response Rate

Objective response was estimated with the greatest precision (Supplementary Table S2). Finotonlimab plus bevacizumab was most effective (odds ratio [OR] 9.51, 2.47–36.74); six of nine credible intervals excluded unity, indicating a response advantage over sorafenib across most combinations. The posterior probability favouring the anti-VEGF-antibody class was 74.8% (τ = 0.31).

3.6 Treatment Rankings

Rankings by SUCRA are presented for all treatments and endpoints in Table 2. Sintilimab plus bevacizumab ranked highest for overall survival (SUCRA 80.7); finotonlimab plus bevacizumab was the most consistent across endpoints, ranking highest for progression-free survival (75.3) and objective response (85.8). Atezolizumab plus cabozantinib ranked lowest for overall survival (18.5), consistent with its null point estimate. Rankings alternated between anti-VEGF-antibody and TKI regimens at the top of the distribution, mirroring the moderate class-level probabilities, and should be interpreted alongside the wide credible intervals as indicative of relative tendencies rather than definitive orderings. These rankings describe the spread of point estimates across the nine trials we have; they are not estimates of comparative drug-class efficacy, and the confounding set out in Section 4.2 applies to them in full.

Table 2 Treatment rankings by SUCRA.

Treatment	Class	OS	PFS	ORR
Sintilimab + bevacizumab	ICI + anti-VEGF	80.7	62.7	66.5
Finotonlimab + bevacizumab	ICI + anti-VEGF	75.3	75.3	85.8
Camrelizumab + rivoceranib	ICI + TKI	69.3	67.3	68.2
Atezolizumab + bevacizumab	ICI + anti-VEGF	65.3	43.7	39.3
Anlotinib + penpulimab	ICI + TKI	60.0	71.6	42.4
Toripalimab + bevacizumab	ICI + anti-VEGF	47.8	37.1	61.2
Pembrolizumab + lenvatinib	ICI + TKI	47.1	64.1	62.9
Lenvatinib	TKI monotherapy †	23.7	43.8	34.7
Atezolizumab + cabozantinib	ICI + TKI	18.5	29.9	37.0

Surface under the cumulative ranking curve (SUCRA, %) for each regimen across the three efficacy endpoints. Higher values indicate a higher probability of ranking among the more effective treatments. Regimens are ordered by overall-survival SUCRA. Sorafenib, the common reference comparator, is omitted. † Lenvatinib is included as the bridging node and is not a first-line combination regimen. Rankings derive from an indirect star network estimated with wide credible intervals and should be interpreted as indicative of relative position rather than definitive orderings. OS, overall survival; PFS, progression-free survival; ORR, objective response rate. SUCRA values describe the distribution of point estimates across the nine trials in this network and are not estimates of comparative drug-class efficacy; see Sections 3.6 and 4.2.

3.7 Risk of Bias

Of the nine trials, five were judged at low risk of bias overall, and four raised some concerns (Supplementary Table S6). Risk of bias for overall survival was low across all trials, as this objective endpoint is unaffected by knowledge of treatment assignment. The concerns identified were confined to the progression-based endpoints in two trials for which blinded independent central review of response could not be confirmed (HEPATORCH and the finotonlimab plus bevacizumab trial) and to the reporting of overall survival at a pre-planned interim analysis in two further trials (ORIENT-32 and APOLLO).

3.8 Safety

Treatment-related grade ≥3 adverse events were available for all nine trials. At the class level, ICI plus anti-VEGF-antibody regimens carried toxicity comparable to sorafenib (median odds ratio 1.09). In contrast, ICI plus TKI regimens were approximately twice as toxic (median odds ratio 2.10). The posterior probability that the best-tolerated anti-VEGF-antibody regimen was less toxic than the best-tolerated TKI regimen was 77.7% (Supplementary Figure S2; $\tau = 0.34$).

The most severe toxicity was concentrated in specific TKI-based regimens: camrelizumab plus rivoceranib carried the highest odds (OR 3.85, 95% CrI 1.10–12.92, the only interval excluding unity), followed by atezolizumab plus cabozantinib (2.50) (Supplementary Figure S2). The two classes differed more in the pattern of toxicity than in its overall frequency (Figure 3). Grade ≥3 hand-foot syndrome occurred in every TKI-based combination and in none of the four anti-VEGF-antibody combinations, while reaching grade ≥3 in 6.9–12% of patients randomised to sorafenib in those same trials.

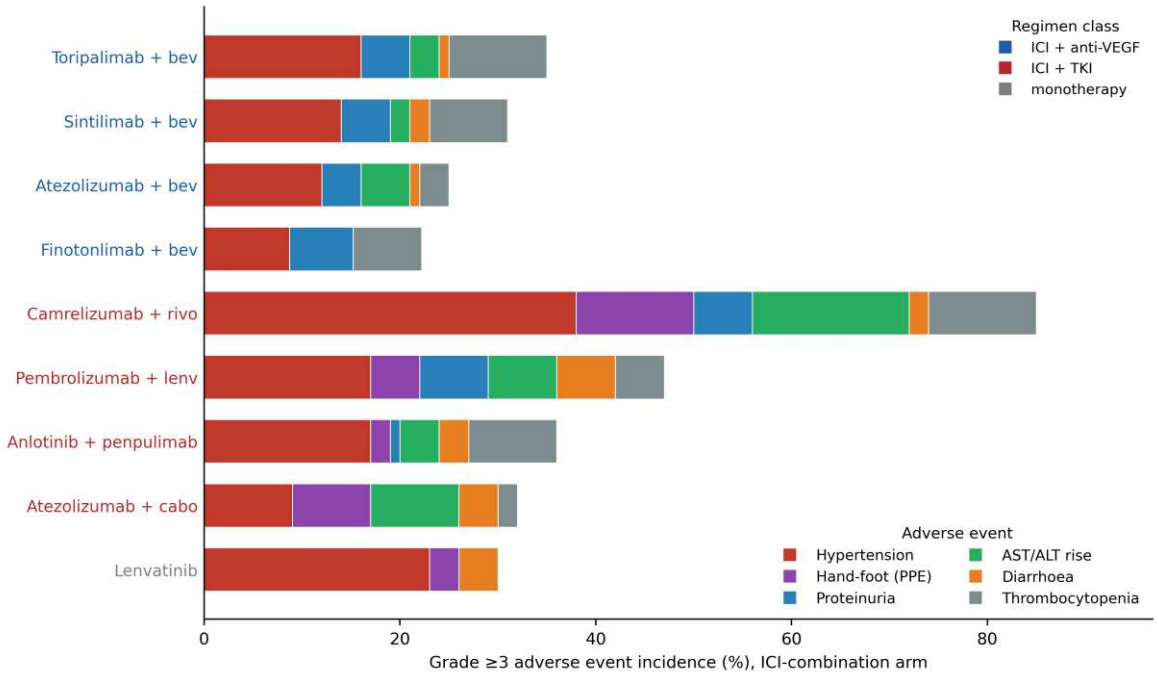


Figure 3 Toxicity signatures: grade ≥3 adverse events by regimen. Incidence of six pre-specified grade ≥3 adverse events in the ICI-combination arm of each trial, extracted from trial safety tables and appendices and grouped by treatment class. Grade ≥3 hand-foot syndrome occurred with every TKI-based regimen (2–12%) and with none of the four ICI plus anti-VEGF-antibody regimens; in each of those four trials it was explicitly reported as absent at grade ≥3 while occurring at grade ≥3 in 6.9–12% of patients randomised to sorafenib in the same trial, so the

absence reflects a within-trial contrast rather than under-reporting. The remaining five adverse events overlapped between classes. Some trials reported treatment-emergent rather than treatment-related events, and for events other than hand-foot syndrome an absent bar may indicate that the event fell below the trial's reporting threshold rather than a confirmed absence. Network estimates of the overall odds of grade ≥3 treatment-related adverse events versus sorafenib are shown in Supplementary Figure S2.

3.9 Effect Modifiers and Their Interdependence

Trial-level meta-regression, interpreted for direction only, showed that greater overall-survival benefit was associated with higher HBV prevalence (slope −0.068 per 10%; p = 0.01), higher macrovascular-invasion/extrahepatic-spread proportion (p = 0.03), and, non-significantly, higher AFP-high proportion (p = 0.21). These modifiers were, however, strongly intercorrelated (pairwise r = 0.81–0.91) and confounded with treatment class and geographic region (HBV–China r = 0.87; AFP–China r = 0.93) and cannot be interpreted as independent effect modifiers (Supplementary Table S7). With nine trials and modifiers correlated at this magnitude the regression is substantially underpowered: it cannot disentangle the contribution of hepatitis B prevalence from that of China-only enrolment, and the slopes above are reported for direction only, as hypothesis-generating signals rather than as estimates of any modifier's independent effect. No inference in this paper rests on their magnitude.

3.10 Sensitivity Analyses

The main overall-survival comparison held up across model choice, prior, and trial mix (Supplementary Table S3). A fixed-effect model raised the probability favouring the anti-VEGF-antibody class to 85.5%. Two alternative predictive distributions, one for drug versus placebo comparisons and one for cause-specific mortality or composite outcomes, gave 74.0% and 73.9%, so the result does not hinge on how the outcome is categorised. Limiting the network to high-HBV trials gave 88.4% (≥80%) and 75.2% (≥70%). Leaving out one trial at a time kept the probability above 50% every time, from 59.6% (without ORIENT-32) to 85.2% (without CARES-310). Since LEAP-002 joins the network only through REFLECT, dropping REFLECT also drops LEAP-002; the remaining seven-trial star, with no bridged evidence, gave 78.2%. The class comparison does not depend on the lenvatinib bridge. But the bridge does make one estimate less precise. Pembrolizumab plus lenvatinib was estimated at HR 0.77 (95% CrI 0.42–1.40), with a posterior standard deviation on the log-hazard scale of 0.30. Re-anchoring the LEAP-002 contrast directly on sorafenib, holding its magnitude and standard error unchanged,

reduced that standard deviation to 0.23, so approximately 41% of the posterior variance of this estimate is attributable to the two-step indirect path rather than to the LEAP-002 data themselves, and the credible interval is approximately 1.32 times wider than it would be under direct anchoring. For comparison, the median posterior standard deviation among the eight regimens anchored directly on sorafenib was 0.23. The pembrolizumab plus lenvatinib estimate therefore carries lower evidential certainty than every other regimen in the network, and is rated accordingly in Supplementary Table S4.

4 Discussion

4.1 Principal findings

This network meta-analysis of nine randomised trials compared first-line combinations of a checkpoint inhibitor with an antiangiogenic agent in advanced HCC. Three regimens improved overall survival over sorafenib with credible intervals excluding unity — sintilimab plus bevacizumab, finotonlimab plus bevacizumab and camrelizumab plus rivoceranib — and six of nine improved objective response. At class level the anti-VEGF-antibody regimens ranked above the TKI regimens for overall survival (75.0%) and objective response (74.8%), with no difference for progression-free survival (45.3%). Our estimates therefore agree in direction with the three published syntheses of this question.

We do not interpret them as class effects, and the remainder of this discussion sets out why we believe the published rankings should not be interpreted that way either.

4.2 Why class attribution is not identifiable in this network

An indirect class comparison carries a causal reading only under transitivity: the trials contributing to each class must be exchangeable with respect to effect modifiers. In this evidence base they are not, and the departure is not marginal. The four trials with the highest hepatitis B virus prevalence (84–94%) were all conducted exclusively in China; globally enrolled trials, including IMbrave150 at 47.9%, sit at the other end of the range. Hepatitis B prevalence correlates with China-only enrolment at r = 0.87 and with high alpha-fetoprotein proportion at r = 0.93, and the pre-specified modifiers correlate with one another at r = 0.81–0.91 (Supplementary Table S7). Treatment class is distributed across the same axis.

The consequence is a question of identifiability rather than of precision. Our exploratory meta-regression found greater survival benefit in trials with more hepatitis B-related disease (slope −0.068 per 10%), a direction that is biologically plausible given the differing immune context of viral and non-viral HCC — though with only nine trials the regression is underpowered, and we do not treat the slope as an effect estimate. But that finding and the class ranking are not two separate results; they are the same result described two ways. A model that explains the advantage by drug class fits these nine trials about as well as a model that explains it by viral aetiology, and the data give little basis for choosing between them. The correlations are high but not perfect, so the confounding is not complete. It is still strong enough that we cannot read the class estimate as a causal effect. Enlarging the network with further trials of the same geographic structure would narrow the credible intervals without altering this.

Our sensitivity analyses illustrate the point rather than resolving it. Across alternative empirical priors, a fixed-effect specification, high-HBV restriction, leave-one-trial-out re-estimation and removal of the lenvatinib bridge, the class probability moved between 73.9% and 85.5% and never fell below 50% (Supplementary Table S3). That robustness is often read as reassurance. Here it is not: a quantity that is confounded at source remains confounded under every reweighting, and the stability of the estimate reflects the stability of the confounding.

4.3 Relation to previously published class rankings

Zhou and colleagues [17] compared the same two classes across six trials and found no significant difference (HR 0.82, 95% CI 0.65–1.04). Chen and colleagues [18] grouped ten trials by viral aetiology and reported greater benefit in hepatitis B-related disease. Zhang and colleagues [19] synthesised fifteen trials in 9,792 patients and likewise ranked PD-(L)1 plus anti-VEGF antibody highest for overall survival. All three report class-level probabilities as estimates of a property of the drugs.

The argument above applies to each of them, because the limitation belongs to the evidence base rather than to any particular analysis. Chen and colleagues tested for an aetiology–treatment interaction and found none ($p = 0.38$); with ten trials and modifiers correlated at this magnitude, a null interaction test has little power to establish independence and should not be read as licensing a causal class interpretation. We make no claim that these analyses were conducted

poorly. We claim that the quantity they report does not have the meaning it is routinely given, and that this is visible from the trial characteristics they themselves tabulate.

4.4 What remains interpretable: within-trial contrasts

Comparisons estimated within randomised groups are untouched by the confounding described above, because the comparator is drawn from the same population as the experimental arm. The clearest example in these data is dermatological toxicity. Grade ≥3 hand-foot syndrome occurred with every TKI-based combination (2–12%) and with none of the four anti-VEGF-antibody combinations, while reaching grade ≥3 in 6.9–12% of patients randomised to sorafenib in those same four trials. The class difference in dermatological toxicity between selective VEGF blockade and broad multikinase inhibition is well established pharmacologically; what the network adds is a demonstration of the contrast between the two contemporary first-line backbones, and an illustration of which class-level statements this evidence base can support and which it cannot.

The same logic qualifies the aggregate safety comparison. At class level the anti-VEGF-antibody regimens carried toxicity comparable to sorafenib (median odds ratio 1.09) and the TKI regimens approximately twice that (2.10), with a 77.7% posterior probability favouring the anti-VEGF-antibody class. Most of the excess arose from two regimens — camrelizumab plus rivoceranib, the only one whose credible interval excluded unity, and atezolizumab plus cabozantinib — while anlotinib plus penpulimab was as well tolerated as the antibody-based regimens despite its TKI backbone. Tolerability tracked the individual agent more closely than the class.

4.5 The atezolizumab plus bevacizumab ranking

Atezolizumab plus bevacizumab, the IMbrave150 standard of care, ranked below several newer regimens [6]. This reflects the structure of indirect ranking rather than a lesser effect: in a star network a regimen ranks on the magnitude of its point estimate against the shared comparator, so a single newer trial with a larger estimate and wider intervals can outrank a larger, longer-followed trial. It is a further reason to read SUCRA orderings in this literature as descriptions of the available estimates rather than as clinical guidance.

4.6 Strengths and limitations

The analysis treats the two antiangiogenic backbones as mechanistic classes, combines efficacy and safety within one framework, examines toxicity event by event rather than only in aggregate, and verifies every extracted value against the primary publication, with two composite adverse-event rates corrected during verification. Its principal limitation is the one it documents: with no head-to-head trials the network is a star around sorafenib with no closed loops, every active comparison is indirect, and consistency cannot be evaluated statistically. A second limitation is structural and bears on every efficacy estimate in the paper. Because each active treatment appears in exactly one trial, the network is saturated: the between-study heterogeneity parameter τ is not identifiable from the included data at all, and its posterior remains close to whichever prior is supplied. We supplied the Turner empirical predictive distributions [28], derived from an external collection of meta-analyses rather than from these nine trials. Every hazard ratio, odds ratio, credible interval and class probability reported here therefore depends on that external prior. Our sensitivity analyses across alternative predictive distributions (Supplementary Table S3) show only that the result holds whichever of these priors we choose; they do not show that it would hold without one. This is a limitation of the evidence base rather than of the model, since no model can estimate a parameter the data say nothing about. The pembrolizumab plus lenvatinib estimate is the least precise in the network, passing twice through the REFLECT bridge; approximately 41% of its posterior variance arises from that second indirect step rather than from the LEAP-002 data, and it should be read as carrying lower certainty than the regimens anchored directly on sorafenib. Restricting the network to the seven directly controlled trials left the class probability at 78.2%, so the class comparison itself does not depend on the bridge. Most trials were open-label, so progression-based endpoints may carry assessment bias where blinded central review could not be confirmed; adverse-event reporting varied between treatment-related and treatment-emergent; and with nine trials, publication bias could not be assessed formally. Sex was reported as a baseline characteristic in all included trials, but outcome data were not available disaggregated by sex, so sex- and gender-based analysis of treatment effect was not possible. Finally, the analysis is confined to first-line therapy; the separate questions of optimal sequencing after progression and of checkpoint-inhibitor rechallenge, where the evidence remains observational, lie outside its scope [31].

5 Conclusions

Among first-line ICI plus antiangiogenic combinations in advanced hepatocellular carcinoma, three regimens improved overall survival over sorafenib and six of nine improved objective response. At class level the anti-VEGF-antibody regimens held a modest, non-decisive advantage for overall survival and objective response and none for progression-free survival.

We do not interpret this as a class effect, and we suggest that the published rankings should not be interpreted as class effects either. Treatment class, viral aetiology and geographic composition vary together across the nine available trials, and aggregate data of this structure cannot separate them; the resulting rankings describe the evidence base rather than a property of the drugs. Contrasts estimated within randomised comparisons, such as the dermatological toxicity difference between the two backbones, remain interpretable and can inform regimen selection and supportive-care planning. Consistent with this, the class-level indirect comparison was rated very low certainty on GRADE-NMA, downgraded for unresolvable confounding by trial population characteristics. Resolving the efficacy question will require individual-patient-data analysis or head-to-head randomisation.

Declarations

Acknowledgements. None.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval. Not applicable. This study is a systematic review and network meta-analysis of published aggregate trial data and did not involve new studies of humans or animals.

Consent to participate. Not applicable.

Consent for publication. Not applicable.

Availability of data and material. All data analysed were extracted from published reports of the included randomised trials, which are cited in the reference list. No new patient-level data were generated. The protocol is registered with PROSPERO (CRD420261353528). The extracted trial-level dataset and verification log are provided in the Supplementary Material.

Code availability. The model specification and analysis code are provided in the Supplementary Material.

Author contributions. Burkhoniddin Tillashaykhov: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft. Iroda Ikromova: Investigation, Data curation, Validation, Writing – review and editing. Dilnavoz Berdiyeva: Conceptualization, Supervision, Project administration, Writing – review and editing. All authors read and approved the final manuscript.

Use of generative artificial intelligence. Generative artificial intelligence was not used to design the study, screen or select studies, extract data, generate or analyse data, or produce figures. All statistical analyses were performed by the authors in PyMC and all extracted values were verified by the authors against the primary trial publications. A large language model (Claude, Anthropic) was used to assist with language editing and manuscript formatting. The authors reviewed and edited all output and take full responsibility for the content.

Prior presentation. An earlier version of this analysis was presented as a poster at ESMO GI 2026 (Munich; Poster #743). No full-length article arising from that presentation has been published or submitted elsewhere.

Supplementary Material

The following supplementary material is provided as a single Electronic Supplementary Material file.

Supplementary Methods. Heterogeneity prior specification and model code.

Fig. S1. PRISMA-NMA flow diagram.

Fig. S2. Grade ≥3 treatment-related adverse events: odds ratios versus sorafenib.

Table S1. Baseline characteristics of included trials.

Table S2. Network estimates for progression-free survival and objective response rate.

Table S3. Sensitivity and robustness analyses of the overall-survival class comparison.

Table S4. GRADE certainty of evidence for overall survival.

Table S5. League table of all pairwise overall-survival comparisons.

Table S6. Risk-of-bias assessment (Cochrane RoB 2), domain-level judgements.

Table S7. Effect-modifier correlations and meta-regression.

Table S8. GRADE-NMA certainty of evidence for the class-level indirect comparison (ICI plus anti-VEGF antibody versus ICI plus VEGFR-TKI).

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–263. https://doi.org/10.3322/caac.21834

2. Wang QB, Luo WL, Li YK, Li J, Yang ZS, Zhao K, Lakang Y, Liang YB, Chen XM, Zuo JX, et al. Tumor compression of the hepatic or portal vein predicts the presence of microvascular invasion and satellite nodules in hepatocellular carcinoma: a retrospective study. J Hepatocell Carcinoma. 2025;12:2055–2067. https://doi.org/10.2147/JHC.S544589

3. Luo WL, Wang QB, Li YK, Liang YB, Li J, Chen XM, Lakang Y, Yang ZS, Zuo JX, Wang W, et al. Impact of middle hepatic vein resection during hemihepatectomy on surgical outcomes and long-term prognosis in hepatocellular carcinoma: a retrospective study. J Hepatocell Carcinoma. 2025;12:2681–2692. https://doi.org/10.2147/JHC.S556306

4. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, de Oliveira AC, Santoro A, Raoul JL, Forner A, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med. 2008;359:378–390. https://doi.org/10.1056/NEJMoa0708857

5. Kudo M, Finn RS, Qin S, Han KH, Ikeda K, Piscaglia F, Baron A, Park JW, Han G, Jassem J, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. Lancet. 2018;391:1163–1173. https://doi.org/10.1016/S0140-6736(18)30207-1

6. Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, Kudo M, Breder V, Merle P, Kaseb AO, et al. Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. N Engl J Med. 2020;382:1894–1905. https://doi.org/10.1056/NEJMoa1915745

7. Cheng AL, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, Lim HY, Kudo M, Breder V, Merle P, et al. Updated efficacy and safety data from IMbrave150: atezolizumab plus bevacizumab vs. sorafenib for unresectable hepatocellular carcinoma. J Hepatol. 2022;76:862–873. https://doi.org/10.1016/j.jhep.2021.11.030

8. Ren Z, Xu J, Bai Y, Xu A, Cang S, Du C, Li Q, Lu Y, Chen Y, Guo Y, et al. Sintilimab plus a bevacizumab biosimilar (IBI305) versus sorafenib in unresectable hepatocellular carcinoma (ORIENT-32): a randomised, open-label, phase 2–3 study. Lancet Oncol. 2021;22:977–990. https://doi.org/10.1016/S1470-2045(21)00252-7

9. Kelley RK, Rimassa L, Cheng AL, Kaseb A, Qin S, Zhu AX, Chan SL, Melkadze T, Sukeepaisarnjaroen W, Breder V, et al. Cabozantinib plus atezolizumab versus sorafenib for advanced hepatocellular carcinoma (COSMIC-312): a multicentre, open-label, randomised, phase 3 trial. Lancet Oncol. 2022;23:995–1008. https://doi.org/10.1016/S1470-2045(22)00326-6

10. Llovet JM, Kudo M, Merle P, Meyer T, Qin S, Ikeda M, Xu R, Edeline J, Ryoo BY, Ren Z, et al. Lenvatinib plus pembrolizumab versus lenvatinib plus placebo for advanced hepatocellular carcinoma (LEAP-002): a randomised, double-blind, phase 3 trial. Lancet Oncol. 2023;24:1399–1410. https://doi.org/10.1016/S1470-2045(23)00469-2

11. Yau T, Kaseb A, Cheng AL, Qin S, Zhu AX, Chan SL, Melkadze T, Sukeepaisarnjaroen W, Breder V, Verset G, et al. Cabozantinib plus atezolizumab versus sorafenib for advanced hepatocellular carcinoma (COSMIC-312): final results of a randomised phase 3 study. Lancet Gastroenterol Hepatol. 2024;9:310–322. https://doi.org/10.1016/S2468-1253(23)00454-5

12. Shi Y, Han G, Zhou J, Shi X, Jia W, Cheng Y, Jin Y, Hua X, Wen T, Wu J, et al. Toripalimab plus bevacizumab versus sorafenib as first-line treatment for advanced hepatocellular carcinoma (HEPATORCH): a randomised, open-label, phase 3 trial. Lancet Gastroenterol Hepatol. 2025;10:658–670. https://doi.org/10.1016/S2468-1253(25)00059-7

13. Zhao C, Zhang Y, Wang G, Zheng J, Chen W, Lu Z, Zhuang L, Gu S, Han L, Zheng Z, et al. Finotonlimab (PD-1 inhibitor) plus bevacizumab (bevacizumab biosimilar) as first-tier therapy for late-stage hepatocellular carcinoma: a randomized phase 2/3 trial. Signal Transduct Target Ther. 2025;10:249. https://doi.org/10.1038/s41392-025-02333-5

14. Zhou J, Bai L, Luo J, Bai Y, Pan Y, Yang X, Gao Y, Shi R, Zhang W, Zheng J, et al. Anlotinib plus penpulimab versus sorafenib in the first-line treatment of unresectable hepatocellular carcinoma (APOLLO): a randomised, controlled, phase 3 trial. Lancet Oncol. 2025;26:719–731. https://doi.org/10.1016/S1470-2045(25)00190-1

15. Qin S, Gu S, Chan SL, Bai Y, Ren Z, Lin X, Chen Z, Jia W, Jin Y, Guo Y, et al. Camrelizumab plus rivoceranib versus sorafenib as first-line therapy for unresectable hepatocellular carcinoma (CARES-310): final analysis of a randomised, open-label, international, phase 3 study. Lancet Oncol. 2025;26:1598–1611. https://doi.org/10.1016/S1470-2045(25)00543-1

16. Zhou J, Sun H, Wang Z, Cong W, Zeng M, Zhou W, Liu L, Wen T, Kuang M, Zhang B, et al. China Liver Cancer Guidelines for the Diagnosis and Treatment of Hepatocellular Carcinoma (2024 Edition). Liver Cancer. 2025;14:779–835. https://doi.org/10.1159/000546574

17. Zhou Y, Li J, Ying J. Anti-PD-1/L1 antibody plus anti-VEGF antibody vs. plus VEGFR-targeted TKI as first-line therapy for unresectable hepatocellular carcinoma: a network meta-analysis. Explor Target Antitumor Ther. 2024;5:568–580. https://doi.org/10.37349/etat.2024.00236

18. Chen W, Chen B, Hu C, Wei Q, Zhang L, Fu W. First-line immunotherapy for advanced hepatocellular carcinoma: a network meta-analysis of randomized trials with overall and HBV/HCV-stratified efficacy and safety. Front Immunol. 2026;17:1693251. https://doi.org/10.3389/fimmu.2026.1693251

19. Zhang Q, Hu Y, Ge Y, Wang R. A dual-level Bayesian network meta-analysis of first-line immunotherapy in advanced hepatocellular carcinoma: strategy-level and regimen-level evidence. Front Immunol. 2026;17:1907830. https://doi.org/10.3389/fimmu.2026.1907830

20. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Med. 2013;11:159. https://doi.org/10.1186/1741-7015-11-159

21. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med. 2013;159:130–137. https://doi.org/10.7326/0003-4819-159-2-201307160-00008

22. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012;3:80–97. https://doi.org/10.1002/jrsm.1037

23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71

24. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JPA, Straus S, Thorlund K, Jansen JP, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015;162:777–784. https://doi.org/10.7326/M14-2385

25. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Mak. 2013;33:607–617. https://doi.org/10.1177/0272989X12458724

26. Abril-Pla O, Andreani V, Carroll C, Dong L, Fonnesbeck CJ, Kochurov M, Kumar R, Lao J, Luhmann CC, Martin OA, et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ Comput Sci. 2023;9:e1516. https://doi.org/10.7717/peerj-cs.1516

652 27. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting
653 results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011;64:163–
654 171. https://doi.org/10.1016/j.jclinepi.2010.03.016

655 28. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT. Predictive distributions for between-
656 study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat Med.
657 2015;34:984–998. https://doi.org/10.1002/sim.6381

658 29. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett
659 MS, Eldridge SM, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ.
660 2019;366:l4898. https://doi.org/10.1136/bmj.l4898

661 30. Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, Kessels AG,
662 Guyatt GH. A GRADE Working Group approach for rating the quality of treatment effect estimates from
663 network meta-analysis. BMJ. 2014;349:g5630. https://doi.org/10.1136/bmj.g5630

664 31. Li J, Liang YB, Wang QB, Luo WL, Chen XM, Lakang Y, Yang ZS, Zuo JX, Li YK, Li Z, et al.
665 Rechallenge with immune checkpoint inhibitors in patients with hepatocellular carcinoma: a narrative
666 review. Liver Cancer. 2025;15:611–630. https://doi.org/10.1159/000549355

Supplementary Files

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

- BMCGastroSupplementaryMaterialR1.pdf