Research article

JHEP|Reports

Hepatocellular carcinoma risk scores for non-viral liver disease: A systematic review and meta-analysis

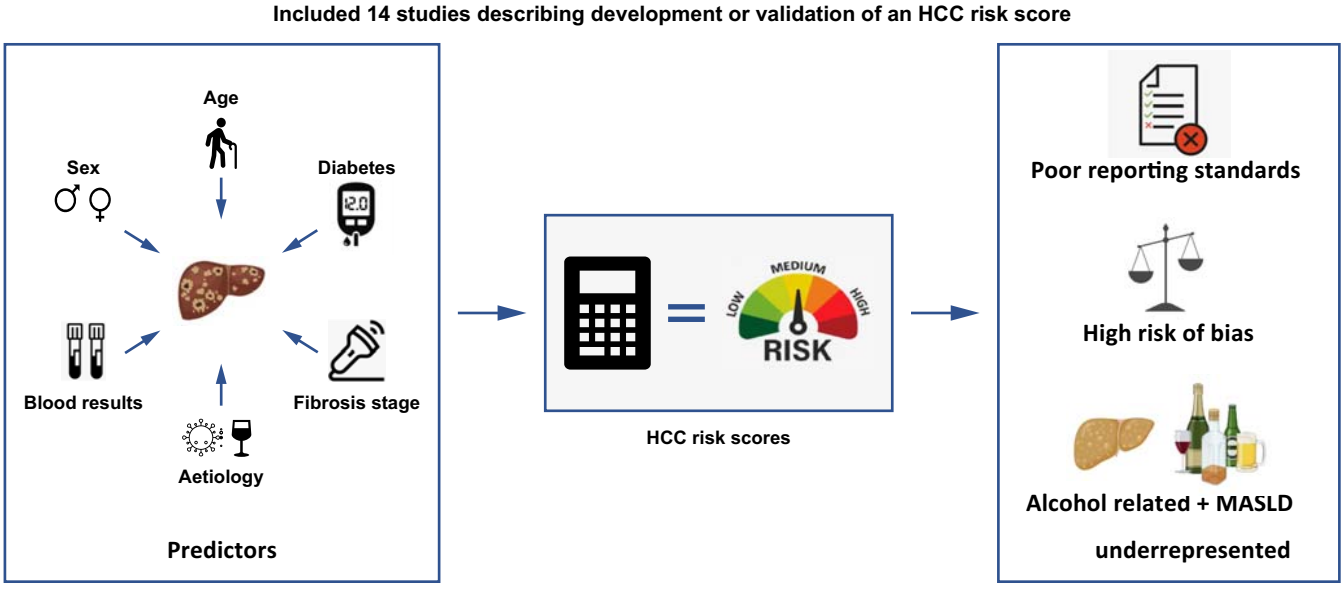
Authors

Laura Burke, Alexander Hinkson, Vincent Haghnejad, Rebecca Jones, Richard Parker, Ian A. Rowe

Correspondence

laura.burke1@nhs.net (L. Burke).

Graphical abstract



Highlights:

- The current literature on HCC prediction models for patients with non-viral liver disease is limited.
- Patients with ALD and MASLD are underrepresented in HCC risk model development studies.
- The competing risks of liver decompensation and death are not accounted for in most HCC prediction model development studies.
- This, along with a high risk of bias, limits the clinical utility of current HCC prediction models in non-viral liver disease.

Impact and implications:

The recent EASL policy statement emphasises the potential of risk-based surveillance to reduce both hepatocellular carcinoma (HCC)-related deaths and surveillance costs. This study addresses the gap in understanding the performance of current HCC risk models in patients with non-viral liver diseases, reflecting the epidemiological landscape of liver disease in Western countries. In our review of these models we identified several key concerns regarding reporting standards and risk of bias and confirmed that patients with alcohol-related liver disease and metabolic dysfunction-associated steatotic liver disease are underrepresented in model development and validation cohorts. Additionally, most models fail to account for the significant risk of competing events, leading to potential overestimation of true HCC risk. This study highlights these critical issues that may hinder the implementation of risk models in clinical practice and offers key recommendations for future model development studies.

https://doi.org/10.1016/j.jhepr.2024.101227

Research article

JHEP|Reports

Hepatocellular carcinoma risk scores for non-viral liver disease: A systematic review and meta-analysis

Laura Burke[1,2,*], Alexander Hinkson[1], Vincent Haghnejad[2,3], Rebecca Jones[1], Richard Parker[1], Ian A. Rowe[1,2]

JHEP Reports **2025**. vol. 7 | 1–12

Background & Aims: Hepatocellular carcinoma (HCC) risk prediction models may provide a more personalised approach to surveillance for HCC among patients with cirrhosis. This systematic review aims to summarise the performance of HCC prediction models in patients with non-viral chronic liver disease.

Method: The study was prospectively registered with PROSPERO (ID: CRD42022370078) and reported in accordance with PRISMA guidelines. MEDLINE and Embase databases were searched using a validated search filter for prediction model studies. Two reviewers independently assessed studies for inclusion and risk of bias. Measures of model performance (discrimination and calibration) to assess the risk of HCC at specified time points were identified. A random effects meta-analysis was performed on a subset of studies that reported performance of the same model.

Results: A total of 7,854 studies were identified. After review, 14 studies with a total of 94,014 participants were included; 45% of patients had viral hepatitis, 27% ALD (alcohol-related liver disease) and 19% MASLD (metabolic dysfunction-associated steatotic liver disease). Follow-up ranged from 15.1–138 months. Only one model was developed using a competing risk approach. Age (7 models) and sex (6 models) were the most frequently included predictors. Model discrimination (AUROC or c-statistic) ranged from 0.61–0.947. Only the 'aMAP' score (age, male sex, albumin, bilirubin, and platelets) had sufficient external validation for quantitative analysis, with a pooled c-statistic of 0.81 (95% CI 0.80–0.83). Calibration was reported in only 9 of 14 studies. All studies were rated at high risk of bias.

Conclusion: Studies describing risk prediction of HCC in non-viral chronic liver disease are poorly reported, have a high risk of bias and do not account for competing risk events. Patients with ALD and MASLD are underrepresented in development and validation cohorts. These factors remain barriers to the clinical utility and uptake of HCC risk models for those with the most common liver diseases.

Introduction

The incidence of hepatocellular carcinoma (HCC) across Europe is increasing.[1] It is now the fifth most common cancer and the second leading cause of cancer-related deaths globally, with a predicted incidence of over 1 million cases by 2025.[2] HCC has among the worst 5-year survival probabilities of any cancer, and, unlike most cancers, mortality related to HCC has increased in recent years.[3] If HCC is detected early, when curative treatments are possible, 5-year survival can exceed 70%.[4] Yet, despite surveillance programmes and treatment advances, the average survival time is currently less than a year.[5]

Over 90% of HCC cases arise in individuals with chronic liver disease (CLD), with cirrhosis being the primary risk factor conveying an annual incidence of greater than 1%.[6] Metabolic dysfunction-associated steatotic liver disease (MASLD) has become the fastest growing liver disease associated with HCC in the West, while alcohol-related liver disease (ALD) is now associated with up to 30% of HCCs world-wide, rising to 60% in certain regions.[7] All current international guidelines recommend biannual abdominal ultrasound surveillance (with or without alpha-fetoprotein) for HCC for patients with cirrhosis.[8–10] However, individual annual risk of HCC varies considerably; from less than 0.2% to over 7%, depending upon multiple risk factors including aetiology, age and sex.[11] This variability raises concerns around the efficacy of a 'one size fits all' approach to screening.[12] Assuming uniform HCC risk across all patients risks harmful under or over estimation of HCC risk for each individual.[13] Although widely accepted as the primary HCC screening tool, ultrasound exhibits suboptimal accuracy in detecting early HCC in cirrhosis, especially in patients with ALD and MASLD cirrhosis due to challenges related to obesity and steatosis.[14] There is therefore a compelling argument that those at the highest risk for developing HCC should be identified and specifically targeted with more effective diagnostic tests.

Given these concerns around current surveillance programme implementation and individual risk variation, there is growing recognition that a more targeted approach is needed. Risk prediction models have the potential to enhance

* Corresponding author. Address: Leeds Liver Unit, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK.
E-mail address: laura.burke1@nhs.net (L. Burke).
https://doi.org/10.1016/j.jhepr.2024.101227

efficiency in HCC surveillance and could enable resources to be focussed on patients who stand to benefit the most from surveillance.[13,15] This approach relies on the availability of appropriate prediction models that can accurately stratify patients. The majority of risk models have been developed from Asian cohorts and focus on patients with chronic HBV[16–19] and HCV.[20,21] This limits their applicability in Europe and North America where MASLD and alcohol aetiologies predominate. Models focussing on HCC risk in 'non-viral' CLD have yet to be integrated into clinical practice and the accuracy of the risk scores have not been assessed. In this study we aimed to systematically identify, collate and critically appraise the performance of HCC prediction models developed when including data from patients with non-viral CLD.

Materials and methods

The systematic review protocol was prospectively registered with PROSPERO (International Prospective Register of Systemic Reviews - ID: CRD42022370078) and was reported in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines.[22]

Search strategy and study selection

A systematic literature search was completed using the OVID platforms MEDLINE (1946 to October 2022), Embase (1947 to October 2022) and the Cochrane library electronic database. Searches were limited to human studies with no language restrictions applied. The search strategy was developed utilising a validated search filter for prediction model studies and a combination of free text terms and medical subject heading searches.[23] The full search strategy is provided in the supplementary materials (Table S1).

Eligibility

Two independent reviewers (LB, AH) performed an initial screen of titles and abstracts. Relevant articles were retrieved for full text evaluation and assessed for eligibility using the predefined inclusion criteria as follows:

- Patients aged 18 years and above.
- Describes the development and/or validation of an original model to predict risk of HCC in patients with CLD using routine clinical tests.
- Study population includes patients with CLD from non-viral aetiologies.
- Presents a measure of performance allowing for assessment of the predictive capacity.

Studies that included individuals with a history of previous HCC or developed models requiring genetic tests or novel biomarkers were excluded. Any disagreements regarding eligibility were resolved by discussion and consensus. A recursive search using the reference list of eligible studies was then performed by hand to identify any additional relevant papers.

Outcome assessment

The model's performance to determine individuals' risk for developing HCC at a specified time point was evaluated. Due to anticipated variability in reporting between individual studies, a range of performance measures were assessed. These included measures of model discrimination (*e.g.* c-statistic, AUROC curve) and model calibration (*e.g.*, calibration plots). The models were also assessed for the inclusion of a competing risks perspective during development.

Data extraction

One reviewer (LB) extracted the relevant data from eligible studies using a standardised proforma derived from the CHARMS checklist[24] and a sample was checked by a second reviewer (IR). The data extraction form included the following study characteristics: Study author, year, time period, geographical location, number of centres included, study design, liver disease aetiologies in cohort, source of patient data, cohort size, follow-up period and HCC incidence in the cohort. Details pertaining to the risk model were extracted including model development method (statistical model or methods employed), risk prediction timing, component variables of model, validation methods and all performance measures of predictive capacity reported.

Quality assessment and risk of bias

Two reviewers (LB, IR) independently assessed the methodological quality of the studies included in the review using the 'transparent reporting of individual prognosis or diagnostic models' (TRIPOD) statement checklist.[25] Any disagreements were resolved by discussion. The studies were then classified into the following types according to the TRIPOD guidelines: development only (1a), development and validation using resampling (1b), random (2a) or non-random (2b) split-sample development and validation, development and validation using separate data,[3] or validation only.[4]

The risk of bias in the studies was assessed by two reviewers (LB, IR), using the PROBAST (Prediction study Risk Of Bias Assessment Tool).[26] This covers four different domains: participant selection, predictors, outcome, and statistical analysis to allow for an overall risk of bias assessment. Further details on how the tool was used are provided in the supplementary materials.

Data synthesis and statistical analysis

Due to the heterogeneity in characteristics of included prediction model studies and the models themselves, we conducted a narrative synthesis. The extracted data was summarised based on the method of model development and predictors included in the final model (allowing for the synthesis of the most used predictors). Events per predictor for each development study was calculated and descriptive statistics relating to model performance including comparisons between the metrics of discrimination (*i.e.* ability to differentiate between patients who develop HCC and those who do not) and metrics of calibration (agreement between the predicted risk of HCC by the model *vs.* the HCC risk observed) are described. A random effects meta-analysis was performed on the subset of studies that reported performance of the same model to generate a pooled c-statistic.

Results

Study characteristics

The search strategy identified 7,854 records. 7,730 studies were excluded following title and abstract review. The full text of the remaining 124 studies were reviewed. Fourteen studies, describing nine original risk models (8 development studies, with 1 study describing 2 models, and 6 external validation studies) met the inclusion criteria and were included in the final qualitative analysis (see Fig. 1).[27–40]

The main characteristics of included studies are shown below in Table 1 and Table 2. Of the 14 studies, 13 were conducted using retrospective data, with 9 of 14 from single-centre cohorts. The majority of studies used data from hospital health records (n = 9) to build or validate the models. Two studies used cohorts from previous trial cohorts,[29,34] one study was based on a cohort from Scientific Registry of Transplant Recipients (127 US liver transplant centres),[38] one from a Veterans Affairs cohort[35] and one from a prospectively collected HCC surveillance cohort.[39]

The studies were conducted over a 24-year period between 1997-2021. The sample size of cohorts ranged from 269 to 34,932 with a total of 94,014 participants across all studies. No study presented an *a priori* sample size calculation. Within the model development studies, derivation cohorts ranged from 442 to 23,243. Half of the studies were conducted in Western populations (n = 3 from the USA, n = 2 from the UK and Australia, n = 1 from Canada and n = 1 from Sweden) and half in Asian populations (Japan n = 2, China n = 1, Korea n = 2, Turkey n = 1, multiple Asian countries n = 1). The median follow-up period across all studies ranged from 15.1 months to 138 months. Overall, a high proportion of study participants (45%, n = 41,870) had viral hepatitis compared with 27% (n = 25,411) with ALD and only 19% (n = 17,640) with MASLD. In five of the eight model development studies, the majority of the study population had viral hepatitis[34,37–40] and only three of the eight development studies had an exclusively non-viral study population.[28,30,35] The proportion of patients with cirrhosis also varied between studies from 6.7% to 100% in the development studies and 25.5% to 100% in the validation studies (see Tables S2 and S3 for study definitions of cirrhosis and HCC).

The number of HCC outcomes in each study varied considerably from 14 HCCs in Liu *et al.*'s aMAP validation study[29] to 1,278 HCC outcomes in Ioannou *et al.* model development study for the 'HCC risk score'[35] (see Table 3). HCC incidence within the study cohorts ranged from 2.57% in the Fan *et al.* aMAP derivation cohort[34] to 14.6% in Zhang *et al.* Toronto HCC risk index (THRI) validation study.[36] All development studies used regression modelling methodology to build their model. The Cox proportional hazards model was used in seven of the eight studies with only one study applying a competing risk perspective using the Fine-Gray sub distribution hazards model.[28] Singal *et al.* developed an additional machine learning model and reported superior performance compared with their regression model based on the same cohort.[39]

A range of two to seven predictors (median = 4) were used in the final models with 15 different predictors used across the eight models (see Table 4 and Fig. S1 for summary of

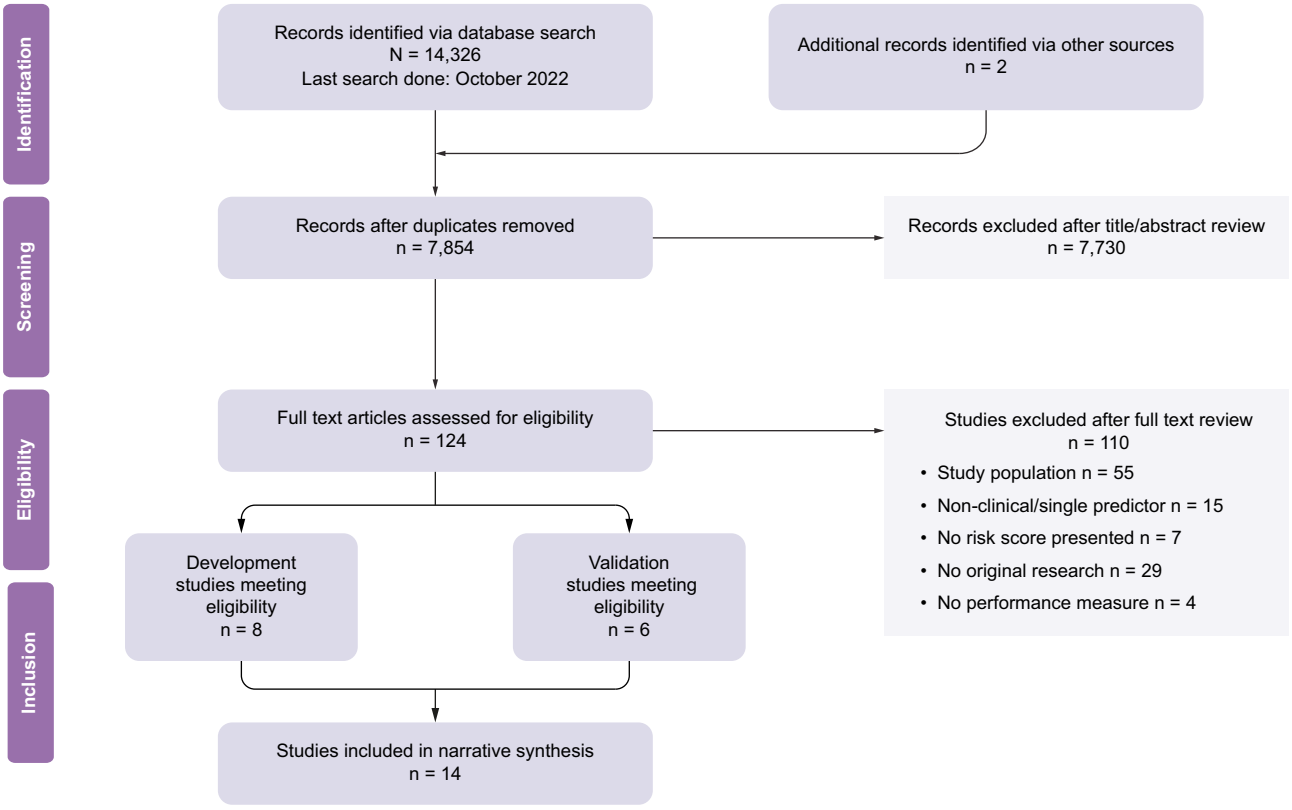


Fig. 1. PRISMA flowchart of study selection process.

Table 1. Characteristics of included HCC risk prediction model development studies.

Study	Name of score	Study region	Study period	Study design	Source of cohort	Aetiology of liver disease (%)	Treatment status of patients with HBV/HCV	Sample size (% cirrhosis)	Median follow-up (months)
Singal 2013[39]	UM model + Machine learning model	USA	2004-2006	Prospective, single-centre	Prospective HCC surveillance cohort	HBV (4.1) HCV (47.3) ALD (14.5) Cryptogenic (19.2) Other (14.9) External validation: HCV (100)	HBV: not reported HCV: not reported HCV: not reported	Total: 1,492 (100) Derivation: 442 (100) External validation: 1,050 (41)	42
Flemming 2014[38]	ADRESS-HCC	USA	2002-2011	Retrospective, multicentre	Scientific Registry of Transplant Recipients	HBV (2.8) HCV (45.7) MASLD (17.5) ALD (18.3) Immune (14.4) Other (1.4) External Validation: HCV (100)	HBV: not reported HCV: not reported HCV: not reported	Total: 34,932 (100) Derivation: 17,124 (100) Validation: 17,808 (100) External validation: 426 (100)	15.12
Aoki 2017[37]	VFMAP	Japan	2008-2013	Retrospective, single-centre	Hospital Health Records	HBV (14.6) HCV (35.1) HCV+HBV (0.6) Non-viral (49.7)	HBV: not reported HCV: not reported	1,808 (12.8)	51.6
Sharma 2017[40]	THRI	Canada	2000-2009	Retrospective, single-centre	Hospital Health Records	HBV[19] HCV (42.3) MASLD (5.4) ALD[11] Immune (12.5) Other (9.7) External validation: HBV (22.1) HCV (42.2) Immune (35.7)	HBV: mixed (76% on treatment) HCV: mixed (22% SVR) HBV: not reported HC: mixed (37.7% SVR)	Total: 3,223 (100) Derivation: 2,079 (100) External validation: 1,144 (100)	64.8
Ioannou 2019[35]	HCC risk score	USA	2012-2018	Retrospective, multicentre	Veterans Affairs Cohort	MASLD (30.4) ALD (69.6)	NA	23,243 (100)	44.4
Fan 2020[34]	aMAP	Multiple	1997-2018	Retrospective, multicentre	Search B, REALM, European PAGE-B, 4 x Gilead HBV RCTs, Japanese HCV trial, STOP-HCV, Scottish HCV, Gilead HCV SVR cirrhotic cohort	HBV (75) HCV (20.5) Non-viral (4.14)	HBV: 100% 'treated' HCV: 100 % SVR	Total 17,374 (30) Derivation: 3,688 (19.3) Validation: Asian CHB: 6,890 (15.1) Caucasian CHB: 2,510 (24.5) HCV: 3,566 (75.2) NVH: 720 (26.3)	42.7
Lee JS 2021[30]	-	Korea	2012-2020	Retrospective, multicentre	Hospital Health Records	MASLD (100)	NA	Total: 3,133 (6.7) Derivation: 2,666 (6.4) Validation 467 (9.4)	60.7
Lee 2022[28]	ALICE	Korea	2004-2017	Retrospective, single-centre	Hospital Health Records	ALD (100)	NA	Total: 745 (100) Derivation: 507 (100) Validation: 238 (100)	59

ALD, alcohol-related liver disease; HCC, hepatocellular carcinoma; MASLD, metabolic dysfunction-associated steatotic liver disease; NVH, non-viral hepatitis; SVR, sustained virological response.

Table 2. Characteristics of included HCC risk prediction model validation studies.

Study	Name of score	Study region	Study period	Study design	Source of cohort	Aetiology of liver disease (%)	Treatment status of patients with HBV/HCV	Sample size (% cirrhosis)	Median follow-up (months)
Zhang 2019[36]	THRI	China	2008-2018	Retrospective, single-centre	Hospital Health Records	HBV (35.3) HCV (1.5) MASLD (0.2) ALD (5.7) Immune (10.9) Other (46.1)	HBV: not reported HCV: no SVR	520 (100)	26.86
Demirtas 2020[33]	THRI	Turkey	2008-2018	Retrospective, single-centre	Hospital Health Records	HBV (39.2) HCV (10.2) MASLD (15.1) ALD (7.1) AIH (8.2) Other (20.3)	HBV: not reported HCV: mixed (48.8% SVR)	403 (100)	58.1
Astrom 2021[31]	THRI	Sweden	2004-2017	Retrospective, single-centre	Hospital Health Records	HBV (4.1) HCV (36.5) MASLD (9.6) ALD (38.1) Immune (5.8) Other (7.1)	HBV: not reported HCV: mixed (11.9% SVR)	2,491 (100)	44.4
Curran 2021[32]	aMAP	UK	2013-2014	Retrospective, single-centre	Hospital Health Records	HBV (1.7) HCV (29.7) MASLD (19.3) ALD (54.4) Immune (2.3)	HBV: not reported HCV: not reported	482 (100)	-
Johnson 2022[27]	aMAP	Japan	1998-2021	Retrospective, single-centre	Hospital Health Records	HBV (24.1) HCV (55) 'Other' (21.3)	HBV: not reported HCV:mixed (<1% SVR at baseline, 53% SVR at end of follow-up)	3,473 (25.5)	138
Liu 2022[29]	aMAP	Australia, UK	-	Retrospective, multicentre	Genom ALC study cohort	ALD (100)	NA	269 (100)	41

ALD, alcohol-related liver disease; HCC, hepatocellular carcinoma; MASLD, metabolic dysfunction-associated steatotic liver disease; NVH, non-viral hepatitis; SVR, sustained virological response. D = development study, V = validation study.

Table 3. HCC outcomes and validation methods of included studies.

Study	Risk prediction period	Number of HCC outcomes (% of study cohort)	Events per predictor (development models only)	Validation method	TRIPOD grade*
Singal 2013[39]	5 years	Derivation: 41 (9.2) Validation: 88 (8.3)	4.1	External cohort	3
Flemming 2014[38]	1 year	Derivation: 1,116 (6.5) Validation: 844 (4.7) Ext validation: 29 (6.8)	85.8	External Cohort	3
Aoki 2017[37]	5 years	49 (2.70)	6.1	None	1a
Sharma 2017[40]	5 years, 10 years	Derivation: 226 (10.80) Validation: 107 (9.35)	25.1	External cohort	3
Ioannou 2019[35]	5 years	1278 (5.50)	51.1	Internal split sample	2a
Zhang 2019[36]	5 years	76 (14.60)	-		4
Demirtas 2020[33]	5 years, 10 years	57 (14.10)	-		4
Fan 2020[34]	3-5 year	Derivation: 95 (2.57) Validation: Asian CHB: 148 (2.14) Caucasian CHB: 147 (5.87) HCV cohort: 222 (6.22) NVH cohort: 19 (2.63)	11.9	External cohort	3
Astrom 2021[31]	5 years, 10 years	304 (12.2)	-		4
Lee JS 2021[30]	2 years, 3 years, 5 years	Derivation: 22 (0.8) Validation: 9 (1.92)	0.44	External cohort	3
Curran 2022[32]	5 Year	22 (4.6)	-		4
Johnson 2022[27]	3 years 5 years	445 (12.81)	-		4
Lee 2022[28]	5 years, 10 years	62 (8.30)	3.3	Internal split sample	2a
Liu 2022[29]	3 years 5 years	14 (5.20)	-		4

HCC, hepatocellular carcinoma.
*Types of prediction model studies according to the TRIPOD guidelines. 1a, development only; 1b, development and validation using resampling; 2a, random split-sample development and validation; 2b, non-random split-sample development and validation; 3, development and validation using separate data; and 4, validation study.

predictors used in models). Age (n = 7) and sex (n = 6) were the most frequently included predictors. The final models were variably presented with two studies reporting models as an equation,[34,38] two studies as a point-based model,[37,40] two as a nomogram[28,30] and one a link to a web-based tool.[35] One study did not present any method to calculate HCC risk from their final model.[39]

Five of the risk model development studies reported external validation.[30,34,38–40] Two were internally validated using split-sample methodology (ALICE, HCC Risk score) and one model had no validation performed (VFMAP). All external validation studies included in the review are for the aMAP (n = 3) or the THRI (n = 3) scores. The rule of thumb 'events per predictor' threshold of >20[26] was met by only 3/8 (38%) model development studies,[35,38,40] and '100 outcome events' for validation studies was met by only 2/6 (33%) validation studies.[27,31]

Model performance

The model performance measures are presented in Table 5. All studies reported a measure of model discrimination. The AUROC was reported in seven studies with a range of 0.758–0.947 in derivation cohorts and 0.707–0.83 in validation cohorts. The model for patients with MASLD from Lee JS *et al.* reported the highest AUROC (0.947, 95% CI 0.914–0.980) of the derivation studies and the aMAP model had the highest AUROC in a validation study (0.83, 95% CI 0.78–0.88).[29,30]

It should be noted that that not all studies specified the type of AUROC reported ('standard' *vs.* 'time-to-event specific' AUROC, see Table 5) and therefore some caution is required in the interpretation of these results given the time-to-event nature of the data. The c-index was reported in five of the development studies, with the aMAP score reporting the highest value (range 0.704–0.82). Only two of the validation studies reported a c-index.[27,31] The highest validation c-index (0.85) was seen in the Japanese HCV validation cohort of the aMAP model study, but this model also had the lowest c-index (0.61) when validated in a 'cirrhosis' cohort.[34]

Measures of calibration were reported in 9 of the 13 studies (Calibration plot in 8, Hosmer-Lemeshow statistic in 1). Ioannou *et al.* also reported a calibration slope of 1 in all their model's derivation cohorts.[35] This dropped to 0.913 in the validation cohort for their ALD model and 0.74 for their NAFLD model. Four of the eight development studies and only two of the six validation studies reported model sensitivity and specificity. Model sensitivity ranged from 53.7 to 93.1% with the ADRESS-HCC model displaying the highest sensitivity from those calculated.[38] Specificity of the models showed a much wider range of 26.2 to 92.2%.[34,38]

Only the 'aMAP' score had sufficient external validation for quantitative analysis. Based on a random effects meta-analysis, the aMAP score had a pooled c-statistic of 0.81 (95% CI 0.80-0.83, Fig. 2). None of the models have undergone prospective assessment of clinical or cost effectiveness.

Table 4. Summary of predictors used in each HCC prediction model.

Model/study name	Predictors, Model n presentation	Model development method	Competing risks considered	Age	Sex	Platelets	Albumin	LSM	Aetiology	AFP	Diabetes	Child Pugh Score	Bilirubin	BMI	AST or ALT/AST ratio	Glucose	Race
UM model Singal 2013	2 Not presented	Cox proportional hazard model	No		X					X							
ADRESS- HCC Fleming 2014	6 Equation	Cox proportional hazard model	No	X	X				X		X	X					X
VFMAP Aoki 2017	5 Point-based model	Cox proportional hazard model	No	X	X			X		X						X	
THRI Sharma 2017	4 Point-based model + nomogram	Cox proportional hazard model	No	X	X	X			X								
HCC risk score Ioannou 2019	7 Web-based tool	Cox proportional hazard model	No	X	X	X	X				X			X	X		
aMAP Fan 2020	5 Equation	Cox proportional hazard model	No	X	X	X	X						X				
- Lee JS 2021	4 Nomogram	Cox proportional hazard model	No	X		X		X							X		
ALICE Lee 2022	3 Nomogram	Fine-Gray subdistribution hazards model	Yes	X			X			X							

AFP, alpha-fetoprotein; ALT, alanine aminotransferase; AST, aspartate aminotransferase; HCC, hepatocellular carcinoma; LSM, liver stiffness measurement.

Reporting standards

Overall publication compliance with TRIPOD ranged from 50 to 78.1% with a median score of 62.9%. The 'HCC risk score' study from Ioannou *et al.* had the best compliance with TRIPOD but still did not satisfy reporting standards in several areas including title, abstract contents, missing data and performance measure reporting.[35] Compliance with abstract reporting (0%), complete reporting of missing data (0%), handling of missing data (25%) and predictor handling (28.6%) was particularly poor across all studies. The full TRIPOD checklist results for each study are provided in Table S4.

Risk of bias

A summary of the risk of bias assessment is presented in Table 6 (Results of all signalling questions available in Tables S5 and S6 and summary of domains in Fig. S2). All studies were rated at overall high risk of bias. Within the individual domains, all studies were rated high risk of bias for the analysis signalling questions. This was largely due to insufficient participants with the outcome, handling of missing data, poor reporting of relevant performance measures and lack of consideration of competing risks within the data. Out of the 14 included studies, 10 were deemed to have a high level of concern in relation to their applicability to the review question. This was largely due to the high proportion of study participants with viral hepatitis.

Discussion

In this systematic review we have comprehensively collated and evaluated the performance of nine HCC prediction models developed when including data from patients with non-viral CLD. While most models demonstrated modest to good discriminative ability, the reporting of model calibration was limited, thereby preventing the comprehensive assessment of their true overall performance. It is particularly noteworthy that patients with MASLD and ALD, the two dominant aetiologies for HCC in Western settings, are underrepresented in both development and validation cohorts.

ALD has remained the major aetiology of liver disease in Northern European countries for the past three decades.[41] Only two of the eight development studies and two of the six validation studies had populations with predominant alcohol aetiologies to reflect this.[28,29,32,35] Whilst the aMAP and THRI scores are intended for use across all aetiologies and are the only models with external validation studies, it is important to highlight that they were derived using predominantly (90.5% and 61.3%, respectively) patients with viral hepatitis.[34,40] The current validation studies in ALD populations are limited by the low number of HCC outcomes observed. Furthermore, despite the well-established association of liver disease aetiology with HCC risk, the aMAP model does not include aetiology as a predictor so further validation is required to determine if this impacts its performance in different populations. Whilst the purpose of this review was to assess performance in non-viral populations it is notable that many studies did not report the treatment/virological response status of patients with viral hepatitis in their study cohorts, again introducing uncertainty regarding model performance for specific populations.

Table 5. Performance measures of HCC risk prediction scores.

Study	C-index (95% CI)	AUROC (95% CI)	Calibration measure	Sensitivity/specificity	PPV/NPV (%)
Singal (machine learning) 2013[39] *D*	Derivation: 0.71 (0.63-0.79) Validation: 0.64 (0.60-0.69)	Not provided	HL stat Derivation: p = 0.69 Validation: p = 0.10	At cut-off 3.25: Se: 80.5 Sp: 57.9	Not provided
Singal 2013[39] *D*	Derivation: 0.64 (0.54-0.73) Validation: 0.61 (0.56-0.67)	Not provided	HL stat Derivation: p = 0.69 Validation: p = 0.01	At cut-off 3.25: Se: 53.7 Sp: 66.6	Not provided
Flemming 2014[38] *D*	Derivation: 0.704 (0.688-0.722) Validation 0.691 (0.672-0.712) External validation: Not provided	Not provided	Calibration plot	At cut-off 4.71: Se: 93.1 Sp: 26.2	Not provided
Aoki 2017[37]	Not provided	0.82 (0.76-0.87)*	Not provided	At cut-off VFMAP 3: Se: 88.4	PPV: 18.9 NPV: 98.2
Sharma 2017[40] *D*	Derivation: 0.76 (0.72-0.79) Validation: 0.77	Not provided	Calibration plot	Not provided	Not provided
Ioannou 2019[35] *D*	Derivation: ALD model: 0.755 MASLD model: 0.749 Joint model: 0.757 Validation: ALD model: 0.740 MASLD model: 0.718 Joint model: 0.727	Derivation:* ALD model: 0.764 MASLD model: 0.775 Joint model: 0.761 Validation: ALD model: 0.740 MASLD model: 0.721 Joint model: 0.738	Calibration plot + slope: Derivation: ALD model:1 MASLD model:1Joint model:1 Validation: ALD model: 0.913 MASLD model: 0.74 Joint model: 0.845	Not provided	Not provided
Demirtas 2020[33] V	Not provided	0.75 (0.683-0.817)*	Not provided	At cut-off 239.5: Se: 78.9 Sp: 62.7	Not provided
Fan 2020[34] D	Derivation: 0.82 (0.77-0.86) Validation: CHB: 0.82-0.87 (?) Japanese HCV:0.85 (0.79-0.91) NVH: 0.85 (0.79-0.90) Cirrhosis: 0.61-0.83	Not provided	Calibration plot	Derivation: Cut-off 50 Se: 86.5 Cut-off 60 Sp: 92.2 Validation Cut-off 50 Se: 85.7-100 Cut-off 60 Sp: 56.6-95.8	Derivation: Cut-off 50 NPV: 95 Cut-off 60 PPV: 13.3 Validation Cut-off 50 NPV: 99.3-100 Cut-off 60 PPV: 6.6-15.7
Zhang 2019[36] V	Not provided	0.707 (0.645-0.769)*	Not provided	At cut-off 226: Se: 84.2 Sp: 48.6	PPV: 21.9 NPV: 94.7
Astrom 2021[31] *V*	0.69	Not provided	Calibration in the large: 0.11 Calibration slope: 1.24	Not provided	Not provided
Lee JS 2021[30] D	Not provided	Derivation: 2 year: 0.948 (0.917-0.979) 3 year: 0.947 (0.914-0.980) 5 year: 0.939 (0.900 -0.978) Validation 2 year: 0.777 (0.6.0-0.948) 3 year: 0.781 (0.614-0.948) 5 year: 0.784 (0.619-0.949)	Calibration plot	Not provided	Not provided
Curran 2022[32] V	Not provided	0.722 (0.636-0.808)*	Not provided	Not provided	Not provided
Johnson 2022[27] *V*	0.81 (0.79-0.82)	Not provided	Calibration plot	Not provided	Not provided
Lee 2022[28] D	Not provided	Derivation 0.758	Calibration plot	Not provided	Not provided
Liu 2022[29] V	Not provided	1 year: 0.83 (0.78-0.88) 3 years: 0.80 (0.69-0.91) 5 years: 0.82 (0.72-0.91)	Not provided	Not provided	Not provided

AUROC, area under the receiver-operating characteristic curve; HCC, hepatocellular carcinoma; H-L, Hosmer-Lemeshow C test; NPV, negative predictive value; PPV, positive predictive; Sp, specificity; Se, sensitivity.
*Denotes studies where type of AUROC used not specified. D = development study, V = validation study.

Table 6. Summary of risk of bias and applicability assessment for each PROBAST domain.

	Risk of bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Development studies									
Singal[39]	LOW	UNCLEAR	LOW	HIGH	HIGH	UNCLEAR	LOW	HIGH	HIGH
Flemming[38]	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH
Aoki[37]	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH
Sharma[38]	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW
Ioannou[35]	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH
Fan[34]	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH
Lee JS[30]	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH
Lee[28]	UNCLEAR	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH
Validation studies									
Zhang[36]	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH
Demirtas[33]	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH
Astrom[31]	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW
Johnson[27]	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH
Liu[29]	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW

Curran *et al.* not included as only abstract available.[32]

The models included in this review demonstrated moderate to good discrimination; that is the ability to differentiate between patients who develop HCC and those who do not. A critical question is what level of discrimination should an HCC risk model possess to be deemed clinically useful? In the absence of comprehensive impact studies, this is difficult to determine. The aMAP model exhibited the best discriminatory performance in both derivation and external validation cohorts with a pooled c-statistic of 0.81 (95% CI 0.80-0.83). This model has also performed well in recent HCV cohort validation studies, which is unsurprising given the similarities of this population to the model derivation cohort.[27,42]

The aMAP study also highlights the importance of exploring model performance within different subgroups of a cohort. This study reported a range of discrimination measures when performance of the model in patients with no cirrhosis *vs.* cirrhosis were compared. This is in line with the validation study from Johnson *et al.* who reported that aMAP performs much better in younger patients and in those without cirrhosis. No other studies in this review explored this aspect of subgroup performance.

Calibration of a model provides insight into the level of agreement between predicted and observed HCC outcomes and is crucial in determining the degree of clinical utility. This critical aspect of performance evaluation was lacking in several of the included studies.[29,32,33,36,37] Despite the publication of guidelines aimed at recognising and addressing the issue of reporting deficiencies in prognostic studies, our current review has confirmed that the problem persists.[25,43,44] Yang *et al.* reviewed reporting of HBV HCC risk models based on the TRIPOD statement with many similar findings to our current review. Missing data reporting was poor, data on model calibration was lacking, with deficiencies in statistical methods, use of small data sets and lack of external validation leading to concerns about overfitting.[19] It is therefore not surprising that none of the prediction models in this review are yet imbedded in clinical practice.[42]

In evaluating predictive models for clinical use, it is essential to consider the risk of bias, the applicability assessment and the results of external validation. These aspects, while related, provide complementary information about a model's performance and reliability. The risk of bias examines factors that could introduce systematic errors or distort the true relationship between predictors and outcomes. The applicability assessment and external validation results test the model's generalisability to ensure that the model maintains its predictive performance across varied settings of the intended population, suggesting robustness and wider applicability.

Previous work has highlighted the potential for existing models to overestimate HCC risk, largely due to ignoring competing risk events.[42,45] It is important to recognise that assessing risk and outcomes in patients with HCC is more complex than many other cancers, as survival is not just determined by cancer stage.[46] Cirrhosis severity impacts treatment options but also affects liver-related mortality, which is a major determinant of long-term survival. If the competing risks of non-HCC deaths are not considered, the real cumulative risks of HCC to the individual can be greatly overestimated.[47] Recent work from Innes *et al.* compared a Cox regression model ignoring competing risk events and a Fine-Gray regression model accounting for non-HCC mortality as

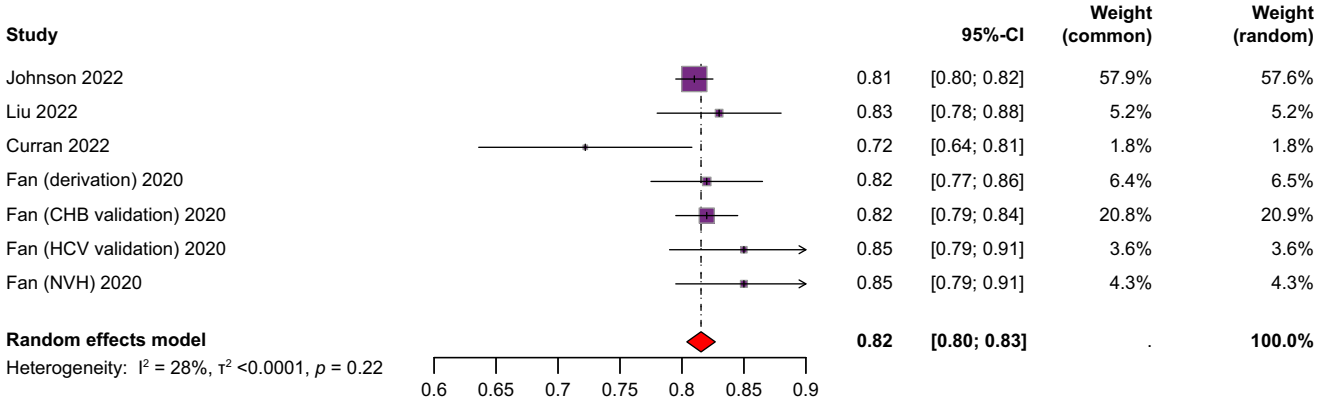


Fig. 2. Forest plot to demonstrate the pooled c-statistic of the aMAP score following random effects meta-analysis.

a competing risk and confirmed that ignoring competing risks leads to overestimation of the probability of developing HCC.[48] Only one model in this review accounted for competing risks during model development. The remainder used a time-to-event regression model that censors all patients with incomplete follow-up, including those with competing events (*e.g.* decompensation and death). A failure to consider these competing risks within the model is likely to lead to overestimation of the absolute risk of HCC and therefore result in misleading prognostication.[46] This is a significant limitation given the high risk of non-HCC events in these patients, especially death from decompensated liver disease in patients with ALD. The issue of competing risks was also demonstrated by Lee *et al.* in their cohort of patients with ALD used to build the 'ALICE' model included in this review.[28] In this study, there more than twice as many non-HCC-related deaths than HCC-related deaths.

The available evidence to date suggests that 'risk-stratified' HCC surveillance strategies targeting high- and intermediate-risk patients are cost-effective and outperform the current surveillance method.[15] This approach to HCC surveillance has also been highlighted in a recent policy statement from EASL, which advocates for risk-based surveillance as a superior strategy to current practice.[49] To our knowledge, this is the first systematic review that focuses on HCC prediction models for non-viral liver disease. Importantly, it has highlighted the issues of performance, applicability and biases of existing models that must be addressed before we can consider implementing risk-based surveillance. Another notable strength of this review is the inclusion of calibration in addition to discrimination when evaluating model performance. Calibration is an aspect frequently overlooked, but poor calibration may make a prediction model clinically useless or even harmful and therefore must be considered if we are to optimise the utility of predictive risk scores for decision-making.

There are some limitations to this review, largely due to the characteristics of the included studies. Our analyses are limited by the low number of available validation studies, quality of the reported data, and high risk of bias in all studies. The differences in study characteristics, populations, and risk factors in the prediction models has resulted in heterogeneity amongst the included studies. In addition, in contrast to the viral hepatitis literature, there are no validation studies that compare multiple risk prediction models in the same cohort to allow for true comparison of performance measures. Finally, there is a wide time range of included studies (1997 to 2021) during which the epidemiological landscape and management of CLD has evolved with a direct impact on both HCC and competing event risks. These limitations highlight the key areas for future work. For a risk-based approach to be accepted by the clinical community we must demonstrate that the best performing models such as aMAP have been assessed in large, multicentric cohorts of patients with adequate follow-up. In our view, the target condition to be predicted by an HCC risk model is the development of HCC before decompensation occurs so that radical treatment can be offered. The competing risk of decompensation occurring before HCC must therefore be accounted for in any future model development and validation studies to ensure accurate prediction. Finally, impact studies are required to quantify the effect of the surveillance risk models on patient outcomes.

In conclusion, current models describing risk prediction of HCC in non-viral CLD are poorly reported and are at high risk of bias. Patients with ALD and MASLD remain underrepresented in development and validation cohorts, and most current models do not account for the greater risk of competing events in patients with these liver diseases.[50] These factors remain barriers to the clinical utility and uptake of HCC risk models in those with the most common liver diseases. Whilst aMAP is arguably emerging as a promising candidate for a universal risk model, further validation in contemporary cohorts where steatotic liver disease predominates is essential before true 'universal' performance can be established.

Affiliations
[1]*Leeds Institute for Medical Research, University of Leeds, Leeds, United Kingdom;* [2]*Leeds Liver Unit, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom;* [3]*Department of Hepatology and Gastroenterology, University Hospital of Nancy, Nancy, France*

Abbreviations

ALD, alcohol-related liver disease; CLD, chronic liver disease; MASLD, metabolic dysfunction-associated steatotic liver disease; THRI, Toronto HCC risk index; TRIPOD, Transparent reporting of individual prognosis or diagnostic.

Financial support

The study was supported by an NIHR research award: Award ID: ALPINE NIHR 155210. VH received a fellowship grant from L'Association Française pour l'Étude du Foie (AFEF).

Conflicts of interest

The authors declare no conflicts of interest that pertain to this work.

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

Authors' contributions

Concept and design of study: LB, IAR; Literature search, data extraction and analysis: LB, AH, IAR; Review of data, drafting and review of manuscript: LB, AH, VH, RJ, RP, IAR.

Data availability statement

The data in this review is based on previously published data with all extracted data available in the manuscript and supplementary materials.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhepr.2024.101227.

References

[1] Rumgay H, Arnold M, Ferlay J, et al. Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol 2022;77(6):1598–1606.
[2] Rahib L, Smith BD, Aizenberg R, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 2014;74(11):2913–2921.
[3] Burton A, Tataru D, Driver RJ, et al. Primary liver cancer in the UK: incidence, incidence-based mortality, and survival by subtype, sex, and nation. JHEP Rep 2021;3(2):100232.
[4] Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. Nat Rev Dis Primers 2021;7(1):6.
[5] Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019;380(15):1450–1462.
[6] Singal AG, Lampertico P, Nahon P. Epidemiology and surveillance for hepatocellular carcinoma: new trends. J Hepatol 2020;72(2):250–261.
[7] Global Burden of Disease Liver Cancer Collaboration, Akinyemiju T, Abera S, Ahmed M, et al. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015. JAMA Oncol 2017;3(12):1683–1691.
[8] Omata M, Cheng AL, Kokudo N, et al. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. Hepatol Int 2017;11(4):317–370.
[9] Marrero JA, Kulik LM, Sirlin CB, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American association for the study of liver diseases. Hepatology 2018;68(2):723–750.
[10] European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2018;69(1):182–236.
[11] Ioannou GN, Green P, Lowy E, et al. Differences in hepatocellular carcinoma risk, predictors and trends over time according to etiology of cirrhosis. PloS one 2018;13(9):e0204412.
[12] Haq MI, Drake TM, Goh TL, et al. Effect of hepatocellular carcinoma surveillance programmes on overall survival in a mixed cirrhotic UK population: a prospective, longitudinal cohort study. J Clin Med 2021;10(13).
[13] Fujiwara N, Friedman SL, Goossens N, et al. Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. J Hepatol 2018;68(3):526–549.
[14] Colli ANT, Miletic D, Giljaca V, et al. Abdominal ultrasound and alpha-foetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. Cochrane Database Syst Rev 2021;(4).
[15] Goossens N, Singal AG, King LY, et al. Cost-effectiveness of risk score-stratified hepatocellular carcinoma screening in patients with cirrhosis. Clin translational Gastroenterol 2017;8(6):e101.
[16] Fung J, Cheung K-S, Wong DK-H, et al. Long-term outcomes and predictive scores for hepatocellular carcinoma and hepatitis B surface antigen seroclearance after hepatitis B e-antigen seroclearance. Hepatology (Baltimore, Md) 2018;68(2):462–472.
[17] Kim JH, Kim YD, Lee M, et al. Modified PAGE-B score predicts the risk of hepatocellular carcinoma in Asians with chronic hepatitis B on antiviral therapy. J Hepatol 2018;69(5):1066–1073.
[18] Yang H-I, Yuen M-F, Chan HL-Y, et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. Lancet Oncol 2011;12(6):568–574.
[19] Yang L, Wang Q, Cui T, et al. Reporting and performance of hepatocellular carcinoma risk prediction models: based on TRIPOD statement and meta-analysis. Can J Gastroenterol Hepatol 2021;2021:9996358.
[20] El-Serag HB, Kanwal F, Davila JA, et al. A new laboratory-based algorithm to predict development of hepatocellular carcinoma in patients with hepatitis C and cirrhosis. Gastroenterology 2014;146(5):1249–12455.e1.
[21] Ganne-Carrie N, Layese R, Bourcier V, et al. Nomogram for individualized prediction of hepatocellular carcinoma occurrence in hepatitis C virus cirrhosis (ANRS CO12 CirVir). Hepatology 2016;64(4):1136–1147.
[22] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
[23] Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012;7(2):e32844.
[24] Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11(10):e1001744.
[25] Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015;13:1.
[26] Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51–58.
[27] Johnson PJ, Innes H, Hughes DM, et al. Evaluation of the aMAP score for hepatocellular carcinoma surveillance: a realistic opportunity to risk stratify. Br J Cancer 2022;127(7):1263–1269.
[28] Lee K, Choi GH, Jang ES, et al. A scoring system for predicting hepatocellular carcinoma risk in alcoholic cirrhosis. Scientific Rep 2022;12(1):1717.
[29] Liu KYT, Masson S, Fateen W, et al. Validation of the aMAP score to predict hepatocellular carcinoma development in a cohort of alcohol-related cirrhosis patients. Liver Cancer Int 2022;3(2):99–104.
[30] Lee JS, Sinn DH, Park SY, et al. Liver stiffness-based risk prediction model for hepatocellular carcinoma in patients with nonalcoholic fatty liver disease. Cancers 2021;13(18).
[31] Astrom H, Ndegwa N, Hagstrom H. External validation of the Toronto hepatocellular carcinoma risk index in a Swedish population. JHEP Rep : innovation Hepatol 2021;3(5):100343.
[32] Curran C, Barclay S. P081YI: validation of the aMAP risk score for the development of hepatocellular cancer in patients with cirrhosis in Glasgow. Digital Liver Cancer Summit Feb 2021:142. https://easl.eu/wp-content/uploads/2021/01/Digital-Liver-Cancer-Summit-2021-Abstract-book.pdf.
[33] Demirtas CO, Gunduz F, Kani HT, et al. External validation of the Toronto hepatocellular carcinoma risk index in Turkish cirrhotic patients. Eur J Gastroenterol Hepatol 2020;32(7):882–888.
[34] Fan R, Papatheodoridis G, Sun J, et al. aMAP risk score predicts hepatocellular carcinoma development in patients with chronic hepatitis. J Hepatol 2020;73(6):1368–1378.
[35] Ioannou GN, Green P, Kerr KF, et al. Models estimating risk of hepatocellular carcinoma in patients with alcohol or NAFLD-related cirrhosis for risk stratification. J Hepatol 2019;71(3):523–533.
[36] Zhang H, Zhu J, Xi L, et al. Validation of the Toronto hepatocellular carcinoma risk index for patients with cirrhosis in China: a retrospective cohort study. World J Surg Oncol 2019;17(1):75.
[37] Aoki T, Iijima H, Tada T, et al. Prediction of development of hepatocellular carcinoma using a new scoring system involving virtual touch quantification in patients with chronic liver diseases. J Gastroenterol 2017;52(1):104–112.
[38] Flemming JA, Yang JD, Vittinghoff E, et al. Risk prediction of hepatocellular carcinoma in patients with cirrhosis: the ADRESS-HCC risk model. Cancer 2014;120(22):3485–3493.

[39] Singal AG, Mukherjee A, Elmunzer BJ, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol 2013;108(11):1723–1730.
[40] Sharma SA, Kowgier M, Hansen BE, et al. Toronto HCC risk index: a validated scoring system to predict 10-year risk of HCC in patients with cirrhosis. J Hepatol 2017.
[41] Williams R, Aithal G, Alexander GJ, et al. Unacceptable failures: the final report of the Lancet Commission into liver disease in the UK. Lancet 2020;395(10219):226–239.
[42] Innes H, Jepsen P, McDonald S, et al. Performance of models to predict hepatocellular carcinoma risk among UK patients with cirrhosis and cured HCV infection. JHEP Rep : innovation Hepatol 2021;3(6):100384.
[43] Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012; 9(5):1–12.
[44] Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162(1):W1–W73.
[45] Innes H, Hamill V, McDonald SA, et al. Comparing predicted probability of hepatocellular carcinoma in patients with cirrhosis with the general population: an opportunity to improve risk communication? Am J Gastroenterol 2022;117(9):1454–1461.
[46] Jepsen P, Vilstrup H, Andersen PK. The clinical course of cirrhosis: the importance of multistate models and competing risks analysis. Hepatology 2015;62(1):292–302.
[47] D'Amico G, Morabito A, D'Amico M, et al. Clinical states of cirrhosis and competing risks. J Hepatol 2018;68(3):563–576.
[48] Innes H, Johnson P, McDonald SA, et al. Competing risk bias in prognostic models predicting hepatocellular carcinoma occurrence: impact on clinical decision-making. Gastro Hep Adv 2022;1(2):129–136.
[49] European Association for the Study of the Liver. EASL Policy Statement: risk-based surveillance for hepatocellular carcinoma among patients with cirrhosis. 2023. Available: https://easl.eu/publication/easl-policy-statement-risk-based/.
[50] Shearer JE, Jones R, Parker R, et al. The natural history of advanced chronic liver disease defined by transient elastography 2023 Mar;21(3):694–703.

Keywords: Carcinoma; Hepatocellular; Liver Cirrhosis; Systematic review; Meta-analysis.
Received 24 June 2024; received in revised form 23 September 2024; accepted 26 September 2024; Available online 9 October 2024