Japanese Journal of Gastroenterology and Hepatology

Research Article

ISSN 2435-1210 |Volume 6

Risk Prediction Models For Liver Cancer In Adults: A Systematic Review

Maomao Cao[1], He Li[1], Dianqin Sun[1], Siyi He[1], Yadi Zheng[1], Zheng Wu[1], Xinyang Yu[1], Lin Lei[2], Ji Peng[2], Jiang Li[1], Ni Li[1], Wanqing Chen[1*]

[1]Office of Cancer Screening, National Cancer Center / National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China.

[2]Department of Cancer Prevention and Control, Shenzhen Center for Chronic Disease Control, Shenzhen 518020, China

[*]Corresponding author:

Wanqing Chen,
Office of Cancer Screening, National Cancer Center/ National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union
Medical College, Beijing 100021, China.
Tel: +86 10 87787039.
E-mail: chenwq@cicams.ac.cn

Received: 01 Apr 2021
Accepted: 22 May 2021
Published: 27 May 2021

Citation:

Wanqing Chen. Risk Prediction Models for Liver Cancer in Adults: A Systematic Review. Japanese J Gstro Hepato. 2021; V6(14): 1-10

Keywords:
Liver cancer; Risk prediction models; Systematic review; PROBAST

1. Abstract

1.1. Background

Prediction of liver cancer risk is beneficial to define high-risk population of liver cancer and guide clinical decisions. We aimed to review and critically appraise the quality of existing risk-prediction models for liver cancer.

1.2. Methods

This systematic review followed the guidelines of CHARMS (Checklist for Critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) and Preferred Reporting Items for Systematic Reviews and Meta (PRISMA). We searched for PubMed, Embase, Web of Science, and the Cochrane Library from inception to July 2020. Prediction model Risk Of Bias Assessment Tool was used to assess the risk of bias of all potential articles. A narrative description was conducted.

1.3. Results

After removal irrespective and duplicated citations, 20 risk prediction publications were finally included. Within the 20 studies, 16 studies performed model derivation and validation process, two publications only conducted developed procedure without validation and two articles were used to validate existing models. Discrimination was expressed as area under curve or C statistic, which was acceptable for most models, ranging from 0.64 to 0.96. Calibration of the predictions model were rarely assessed. All models were graded at high risk of bias. The risk bias of applicability in 13 studies was considered low.

1.4. Conclusions

This systematic review gives an overall review of the prediction risk models for liver cancer, pointing out several methodological issues in their development. No prediction risk models were recommended due to the high risk of bias.

1.5. Systematic review registration: This systematic has been registered in PROSPERO (International Prospective Register of Systemic Review: CRD42020203244).

2. Background

Liver cancer is the sixth most frequently diagnosed cancer and the second largest cause of cancer deaths worldwide, with an estimated 0.91 million cases and 0.83 million deaths in 2020, nearly half occurring in China [1]. The prognosis of patients diagnosed with liver cancer is generally poor, with a 5-year survival rate of less than 20% [2]. Liver cancer screening for individuals at high-risk for liver cancer is beneficial to diagnose as an early stage, and further, prolong survival [3-7]. Risk-based prediction models play an important role in identifying the target population.

More recently, a number of risk prediction models have been developed to predict the risk of liver cancer development [8-11]. Accurate

prediction of liver cancer risk can not only facilitate screening programs but optimize the utilization of healthcare resources. Briefly, a prediction model is a statistical equation based on multiple predictors. A key stage of prediction model research is model development [12]. In this step, several risk factors associated with the development of liver cancer were identified initially, and further, the hazard of each risk factor related to liver cancer was calculated. The secondary stage is model validation. The common methods to validate the accuracy of the prediction models were cross-validation, bootstrap resampling, and external validation in an independent cohort. Discrimination and calibration are used to measure the performance of a prediction model. Several reviews have summarized the risk prediction models for lung cancer [13], colorectal cancer [14], and prostate cancer [15]. However, the quality of liver cancer prediction models remains unknown.

Consequently, a systematic review was conducted to identify and critically appraise published multivariable prediction models for liver cancer. We aimed to describe their characteristics, and understand methodological pitfalls. This study will help to determine future efforts in this field. For potential bias consideration, the Prediction model Risk Of Bias Assessment Tool (PROBAST) was utilized to assess the bias risk of the methodological features of studies [16].

3. Material and Methods

3.1. Study design

This systematic review adhered to the CHARMS (Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) [17] and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta) [18] guidelines (see the PRISMA in the Support information). The protocol for this study was registered on *PROSPERO* database (registration number: CRD42020203244).

3.2. Inclusion and Exclusion criteria

All study reporting on the development, validation, or updating of a multivariable model involving at least two predictors to forecast the risk of the occurrence of liver cancer were included. The detailed description of the inclusion criteria was presented as followed by the PICO (s) (i.e.,: participants, intervention, comparator, and outcomes) criteria:

1. Participants: We included studies that recruited population aged 18 years and older but not prior diagnosed with liver cancer. Study was not considered if the proportion of patients that received any treatment (i.e.: transplant or anti-virus therapy) was over 20%.

2. Intervention: Not applicable.

3. Comparator: Not applicable.

4. Outcomes: The occurrence of liver cancer.

5. Study designs: Study types included retrospective cohort, prospective cohort and case-control designs.

The external validation risk prediction models in another independent dataset were considered original studies but have to provide acceptable measures of model performance.

We excluded the non-original study (commentaries, reviews, editorials, guidelines, case-report, and letters), methodological studies, and conference abstract. Non-human studies, without full text, only logistic regression without a prediction model, failure to report any performance measures were also excluded.

3.3. Literature search

A systematic search for articles in four electronic international databases, including PubMed, Embase, Web of Science, and the Cochrane Library, was performed. All the original English publications from inception until 31 July 2020 were searched without any restriction of countries. A reference list of each eligible article was also independently screened to identify additional studies left out based on the searching strategy. The following terms were used to identify prediction models in PubMed: (risk assessment OR risk prediction OR risk score OR risk calculation OR prediction model OR predict index OR decision rule OR discrimination OR ROC Curve OR calibration OR AUC OR area under the curve OR machine learning OR neural networks computer OR artificial intelligence OR Risk estimation OR Nomogram OR Scoring System OR outcome prediction OR risk classification OR forecasting OR forecast OR decision tree OR predictive score OR validat*). The detailed search strategy can be found in the Supplementary document (Supplementary Panel 1-4). Two researchers (MM C, H L) did the literature search independently.

3.4. Study selection and Data extraction

The selection process was executed as followed based on the predefined eligibility criteria: at the first step, titles and abstract of all papers retrieved by the search strategy were screened for relevance individually, and those irrelevant were discarded. In the second step, two reviewers independently screened full-texts related to our theme. Any disagreement was resolved by discussion or by consulting a third author until consensus was reached.

A standardized extraction form was created by our organizing committee following recommendations of the CHARMS checklist. The same reviewers independently extracted data. Data were extracted based the following items: article information (e.g., author, year of publication, country of publication), source of data, study design (e.g., prospective or case-control), duration of follow-up, participants characteristic (e.g., age, gender, the inclusion and exclusion criteria), sample size (e.g., number of participants, number of patients with the target event), predictors candidate, predictors in the final models, methods for selection of predictors in prediction models, (e.g. cox regression, neural network), handling of categorical and continuous variables, handling of missing data, model performance (e.g., calibration, discrimination, sensitivity, specificity, and classification measures), model presentation, model evaluation (internal and external validation).

3.5. Assessment of study quality and statistical analysis

The quality of the included studies was appraised upon PROBAST, a risk of bias assessment tool. It has four domains: participants, predictors, outcomes, and statistical analysis. Questions are answered with yes, probably yes, probably no, no, and no information, depending on the characteristics of the study. A domain contains at least one question marked as no or probably no, and the overall risk of bias was rated to be at high risk. The overall risk of bias is graded as low risk only all domains were considered low risk. Other circumstance was rated unclear. Applicability was assessed upon the first three domains. Two researchers (MM C, H L) independently assessed the risk of bias, and verified the accuracy by a third author. A narrative descriptive analysis was conducted to summary the basic characteristics of the included studies due to the substantial heterogeneity.

4. Results

4.1. Study selection

After removal of duplicates, a total of 13,612 articles were recruited. Of these, 13,467 records were deleted after titles and abstract screening. The remaining 145 potential studies were further checked for full-text and only 20 studies were included finally. 125 publications were excluded for the following reasons: seven articles were excluded as conference abstract or non-original study; 81 articles were excluded based on the questions related to the models, including not prediction models, only consisting of single predictor, model application or not liver cancer prediction. 40 articles were removed due to the study population (see Figure 1 for a flowchart).

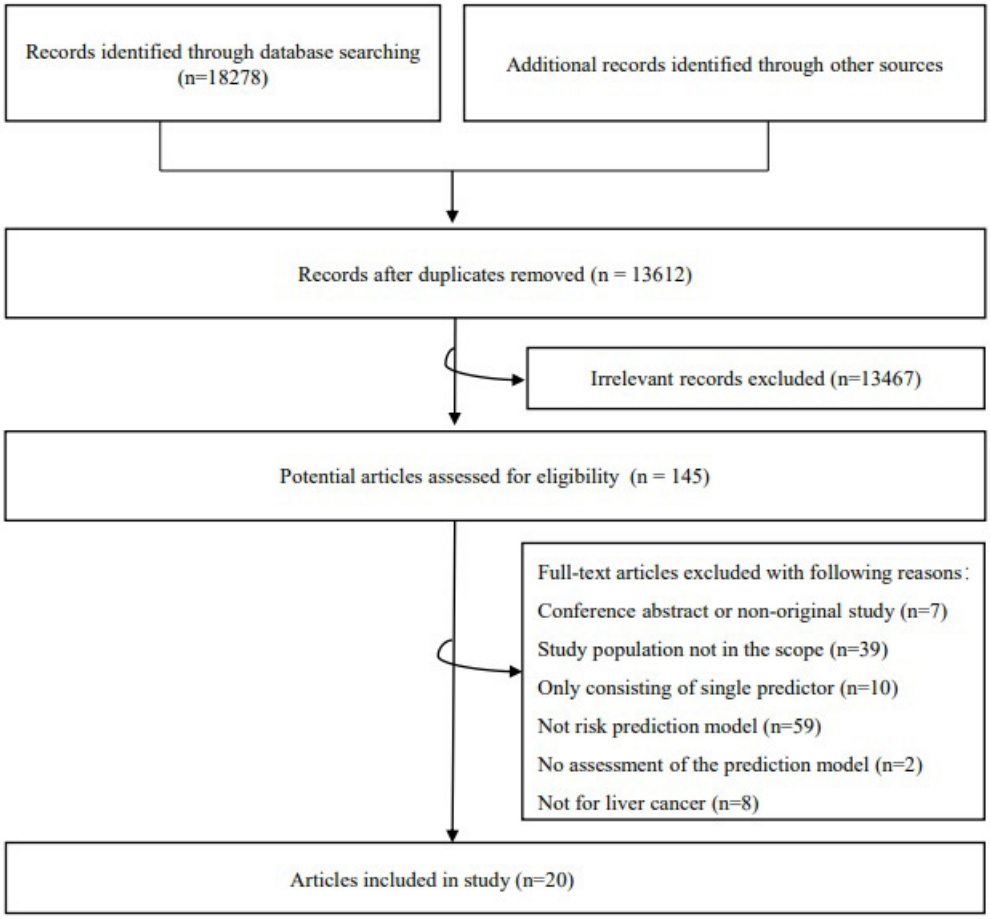

Figure 1: Preferred reporting items for systematic reviews and meta-analyses flowchart for study selection

4.2. Basic characteristics of prediction models

Table 1 and Supplementary Table 1 presents the main characteristics of included studies. All reports were conducted between 2009 and 2020. Most of these studies were conducted in China (N=10, 50%), [8, 11, 19-25] USA (N=3, 15%), [26-29] Korea (N=2, 10%),[30, 31] and Japan (N=2, 10%) [32, 33]. Among all the observational studies, 15 were prospective cohort study and four were retrospective cohort study. And only one study was case-control study. Most were conducted at high risk population, including patients with seropositive hepatitis B surface antigen, chronic hepatitis C virus infection and cirrhosis. Only two studies were initiated based on data from general population [29, 33].

Table 1: Basic characteristics of the prediction models for liver cancer

Author, year	Study design	Country	Modelling method	Predictors in the final model	Follow-up	Sample size
Yuen 2009 [24]	Prospective cohort	Chin a	Cox proportional hazard model	Age, sex, HBV DNA levels, core promoter mutations, cirrhosis	Mean: 76.8 months	820
Wong 2010 [8]	Prospective cohort	China	Cox proportional hazard model	Age, albumin, bilirubin, HBV DNA, cirrhosis	Median: 9.94 (10.53)† years	1,005(424)
Yang 2011 [23]	Prospective cohort	China	Cox proportional hazard model	Age, sex, ALT, HBeAg status, HBV DNA level	Median: 12.0 (7.0)¶ years	3,584(1,505)
Michikawa 2012 [33]	Prospective cohort	Japan	Cox proportional hazard model	Age, sex, alcohol consumption, BMI, diabetes, coffee consumption, HBV, HCV	Mean: 12.6 months	17,654
Wen 2012 [29]	Prospective cohort	USA	Cox proportional hazard model	Age, sex, smoking, alcohol drinking, physical activity, diabetes, AST, ALT, AFP, HBV, HCV	Mean: 8.5 years	130,533
Tseng 2013 [22]	Prospective cohort	China	Cox proportional hazard model	HBV DNA and HBsAg levels	Mean: 14.9 years	2,165
Lee 2013 [20]	Prospective cohort	China	Cox proportional hazard model	Age, sex, HBeAg status, HBV DNA level, ALT, HBsAg level, genotypes C, family history	Person years: 53,551	3,340
Singal 2013 [28]	Prospective cohort	USA	Machine learning and Cox regression model	AST, ALT, the presence of ascites, bilirubin, baseline AFP level, and albumin	Median: 3.5(5.7) years	442 (1,050)
El-Serag 2014 [26]	Prospective cohort	USA	Logistic regression model	AFP, platelets, ALT, interaction terms along with age at time of AFP test	NR	11,721 (5,760)
Flemming 2014 [34]	Prospective cohort	Canada	Cox proportional hazard model	Age, sex, diabetes, race, etiology of cirrhosis, severity of liver dysfunction	Person-years: 36,719 (30,295)	17,124 (17,808)
Lee 2014 [19]	Prospective cohort	China	Cox proportional hazard model	Age, ALT, the ratio of aspirate aminotransferase to ALT, HCV RNA levels, cirrhosis, HCV genotype	Person-years: 14,821(2,265)	975 (572)
Hung 2015 [11]	Prospective cohort	China	Cox proportional hazard model	Age, sex, ALT, CLD, family history of HCC, smoking	Median; 18.8 (18.8) years	8,252(4,125)
Ganne-Carrié 2016 [35]	Prospective cohort	France	Cox proportional hazard model	Age, alcohol intake, platelet count, GGT, absence of SVR	Median 49.0 (54.4) months	720 (360)
Rau 2016 [21]	Case-control	China	Machine learning and logistic regression	Alcoholic cirrhosis, cirrhosis, viral hepatitis, chronic hepatitis, hyperlipidemia.	_____	1,442(618)
Aoki 2017 [32]	Prospective cohort	Japan	Cox proportional hazard model	Age, sex, AFP, VTQ, FPG*	Median: 51.6 months	1,808
Chung 2017 [30]	Retrospective cohort	Republic of Korea	Logistic regression model	Age, sex, cirrhosis, AFP levels and positive US test	Median: 62 (63) months	2,087 (2,088)
Zhang 2019 [25]	Retrospective cohort	China	Logistic regression model	Age, etiology of cirrhosis, sex and platelets	Mean: 32.4 months	520
Ioannou 2019 [27]	Retrospective cohort	USA	Cox proportional hazard model	Age, sex, BMI, diabetes, platelet count, serum albumin, serum AST/ALT ratio	Mean: 3.7 years	23,243
Demirtas 2020 [49]	Retrospective cohort	Turkey	_____	Age, etiology of cirrhosis, sex, and platelet	Mean: 58.1 months	403
Sinn 2020 [31]	Prospective cohort	Korea	Cox proportional hazard model	Age, sex, smoking, diabetes, TC level, ALT	Median: 8.0 (4.6) years	467,206 (91,357)

Abbreviation: HBV, hepatitis B virus; HBeAg, hepatitis B envelope antigen; ALT, alanine aminotransferase; AST, aspirate aminotransferase; BMI, body mass index; AFP, alpha-fetoprotein; HBsAg, hepatitis B surface antigen; CLD, chronic liver disease; SVR, sustained virological response; GGT, gamma-glutamyl transpeptidase; VTQ, Virtual touch quantification; FPG, fasting plasma glucose; TC, total cholesterol.

†: Information for validation set. ¶: Mean follow-up duration

4.3. Follow up, sample size and predictors

The longest median or mean duration of follow-up time was 18.8 years [11]. The sample size of the model derivation varied from 442 to 407,206. Smaller sample size of the validation set was observed, compared to derivation dataset, and the largest sample size in the validation dataset was 91,357 [31]. Every per events (EPVs) ranged from 3.3 [24] to 89.73 [26]. Predictors differed largely from the eligible articles. 17 unique predictor variables were identified among 20 prediction models. The most commonly used predictors for the prediction model were age (n=17; 85%), sex (n=13; 65%), alanine aminotransferase (ALT) (n=7; 35%), alpha-fetoprotein (AFP) (n=5; 25%), cirrhosis (n=5; 25%), platelet count (n=5; 25%), diabetes (n=5; 25%), and HBV DNA (n=5; 25%). Additional information on models could be found at Supplementary Table2.

Table2: Model performance characteristics in the included studies

Studies	Development or validation	Model performance	
		Development	Validation
Yuen 2009 [24]	Development, internal validation	AUC: 0.88, sensitivity, 84.1% (67.7-97.5), specificity: 76.2% (60.8-90.7), PPV: 14.0% (10.0-26.3)，NPV: 98.3% (99.5-99.9)	Sensitivity: 87.9% (74.0-100.0), specificity: 76.2% (73.3-79.1), PPV: 14.6% (12.2-17.2), NPV: 99.3% (98.4-100.0)
Wong 2010 [8]	Development, external validation	Sensitivity: 88.6% (82.5-94.7), NPV: 97.8% (96.6-99.0)	AUC at 5 years: 0.76 (0.66-0.86), sensitivity: 78.3% (74.3-82.2), specificity: 72.8% (68.5-77.2), PPV: 14.2% (10.9-17.5), NPV: 98.3% (89.6-100.0)
Yang 2011 [23]	Development, external validation	AUC at 5 years: 0.80 (0.78-0.82)	AUC: 0.78 (0.76-0.81), calibration: correlation coeffcient (0.99)
Michikawa 2012 [33]	Development, internal validation	C statistic: 0.94	C statistic: 0.94, calibration: overall observed/expected ratio: 1.03 (0.83-1.29)
Wen 2012 [29]	Development, internal validation	AUC: 0.93 (0.93-0.95), calibration: calibration plot	calibration: calibration plot
Tseng 2013 [22]	Development	AUC at 10 years: 0.74 (0.68-0.79)	_____
Lee 2013 [20]	Development, internal validation	AUC at 5 years: 0.89	AUC at 5 years: 0.84
Singal 2013 [28]	Development, external validation	C statistic: 0.64 (0.54-0.73), calibration: HL (p=0.69)	C statistic: 0.61 (056-0.67), calibration: HL test (p<0.001)
El-Serag 2014 [26]	Development, internal validation	PPV: 62%, calibration: HL test (P=0.95)	C statistic: 81.5%, calibration: HL test (p=0.64)
Flemming 2014 [34]	Development, internal, and external validation	C statistic: 0.71 (0.69-0.72), calibration: calibration plot	C statistic: 0.69 (0.67-0.71)
Lee 2014 [19]	Development, external validation	AUC at 5 years: 0.75	AUC: 0.73
Hung 2015 [11]	Development; external validation	C statistic: 0.79 (0.77-0.81), calibration: HL test (P=0.405)	Calibration: HL test (P=0.731)
Ganne-Carrié 2016 [35]	Development, internal validation	AUC at 3 years: 0.72 (0.66-0.77), C statistic: 0.72, calibration: correlation coeffcient (0.91)	AUC: 0.74 (0.64-0.83), calibration: correlation coeffcient (0.86)
Rau 2016 [21]	Development, internal validation	AUC: 0.78 (0.76-0.80); sensitivity: 0.67, specificity: 0.79, PPV: 0.68, NPV: 0.83	_____
Aoki 2017 [32]	Development	AUC: 0.82 (0.76-0.87), sensitivity: 88.4%, specificity: 63.0%, PPV: 18.9%, NPV: 98.2%	_____
Chung 2017 [30]	Development, internal validation	C statistic: 0.96 (0.96-0.96), sensitivity: 78.6% (67.1-87.5), specificity: 96.1% (95.7-96.4), calibration: HL test (p=0.72)	C statistic: 0.94 (0.93-0.94), sensitivity: 67.8% (54.4-79.4), specificity: 95.9 (95.6-96.2), calibration: HL test (p=0.82)
Zhang 2019 [25]	External validation		AUC: 0.71 (0.65-0.77), sensitivity: 0.84, specificity: 0.49
Ioannou 2019 [27]	Development, internal validation	AUC: 0.76, C statistic: 0.76, calibration: calibration slope (1)	AUC: 0.75, C statistic: 0.74, calibration: calibration slope (0.95)
Demirtas,2020 [49]	External validation	_____	AUC: 0.75 (0.68-0.82), sensitivity: 78.9%, specificity: 62.7%
Sinn 2020 [31]	Development, external validation	AUC at 10 years: 0.83 (0.77-0.88), c statistics: 0.80 (0.74-0.87), calibration: correlation coefficient (1).	AUC at 10 years: 0.92 (0.89-0.95), calibration: correlation coefficient (1)
Abbreviation: AUC, Area under curve; PPV, Positive predictive value; NPV, Negative predictive value; HL, Hosmer-Lemeshow.			

4.4. Methodological assessment and missing data

Cox proportional hazards model were used in most studies. One study applied the machine learning and logistic regress model for predicting risk simultaneously [21]. For missing data, most studies have not reported. Complete case analysis [25, 27, 32, 34] or multiple imputation [30] were used to handle missing data.

4.5. Summary of model performance

The model performance measures are presented in Table 2. The most commonly described measure of discriminatory value was AUC, ranging from 0.64 [28] to 0.93 [29] for model development, and 0.73 [19] to 0.92 [31]for model validation. C statistic was reported on nine articles [11, 26 – 28, 30, 31, 33-35] and three articles both described the C statistic and AUC, [27, 31, 35] these C statistic ranged between 0.64 [28] and 0.96 [30]. The discriminatory ability in most studies was considered good (C statistics over 0.7). Nine articles did not report calibration. Internal validation was carried out in ten models [20, 21, 24, 26, 27, 29, 20, 22-35]. Only seven studies conducted external validation.

4.5. Risk of bias for included studies

A summary of the risk of bias assessment of models by domains is presented in table3, and details on each item across domains are presented in Supplementary Table 3. Of the 20 models, 12 studies were defined at low risk of the participants, predictors and outcome domains. However, all models were classified as overall high risk of bias, due to a low number of events per variable, lack of internal validation, less reporting information on missing data and performance measures. Only five studies were assessed at high risk of bias for the participants [21, 22, 25, 28, 30], suggesting that the target population of models has good representatives. Only two of these studies were rated at high risk in terms of predictors [11, 12]. Liver cancer was diagnosed based on published guidelines or sophisticated criteria in 16 studies and thus evaluated at low risk. For applicability, 13 studies were defined at low risk, and seven studies were assessed at high risk [11, 21, 24, 26, 28, 32, 34] which were less potentially applicable to the real setting.

Table 3: Risk of bias assessment based on PROBAST

Studies	Risk of Bias (ROB)				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Michikawa 2012 [33]	Low	Low	Low	High	Low	Low	Low	High	Low
Tseng 2013 [22]	High	Low	Low	High	Low	Low	Low	High	Low
Aoki 2017 [32]	Low	Low	Low	High	High	Low	Low	High	High
Zhang 2019 [25]	High	Low	Low	High	Low	Low	Low	High	Low
Demirtas 2019 [49]	Low	Low	Low	High	Low	Low	Low	High	Low
Yuen 2009 [24]	Low	Low	Low	High	High	Low	Low	High	High
Wong 2010 [8]	Low	Low	Unclear	High	Low	Low	Low	High	Low
Yang 2011 [23]	Low	Low	Low	High	Low	Low	Low	High	Low
Wen 2012 [29]	Low	Low	Low	High	Low	Low	Low	High	Low
Lee 2013 [20]	Low	Unclear	Low	High	Low	Low	Low	High	Low
Singal 2013 [28]	High	Low	Low	High	High	Low	Low	High	High
El-Serag 2014 [26]	Low	Low	High	High	Low	Low	High	High	High
Flemming 2014 [34]	Low	Low	High	High	Low	Low	High	High	High
Lee 2014 [19]	Low	Low	Low	High	Low	Low	Low	High	Low
Hung 2015 [11]	Low	High	Low	High	High	Low	Low	High	High
Rau 2016 [21]	High	High	High	High	High	Low	High	High	High
Ganne-Carrié 2016 [35]	Low	Low	Low	High	Low	Low	Low	High	Low
Chung 2017 [30]	High	Low	Low	High	Low	Low	Low	High	Low
Ioannou 2019 [27]	Low	Low	Low	High	Low	Low	Low	High	Low
Sinn 2020 [31]	Low	Low	Low	High	Low	Low	Low	High	Low

5. Discussion

Risk prediction models for liver cancer have become common in recent years, but the quality and bias haven't been assessed. In this systematic review, we identified and critically appraised 20 studies reporting multivariable prediction models for liver cancer. Most studies were modeling based on high-risk population. Discrimination of models was considered acceptable in most studies but the calibration was less reported. All models were appraised to have a high risk of bias due to a combination of poor reporting and methodological shortcomings. Because of the high risk of bias, no model is recommended to being used in a practical environment.

To minimize bias, a prospective cohort design was applied to the ma-

jority of studies, and the Cox proportional method to predict risk of the development of liver cancer and potential risk factors was used. The strongest advantage of this design was that the potential risk factors were assessed carefully before liver cancer diagnosis. In addition, information on all participants was acquired under the supervision of the researchers, thus, the results were relatively reliable. However, a prospective cohort study is difficult to conduct, which takes at least one or two years. Conversely, a retrospective cohort study is comparatively less costly based on the immediate availability of the data [36]. But the completeness and authenticity of this design was limited. In this study, retrospective cohort studies with logistic regression model were applied in two studies [25, 30]. Case-control design was used in only one study [21], which needed relatively fewer subjects, and thus the accuracy was less reliable.

After the risk of bias evaluation, all models were graded at a high risk of bias for statistical analysis. This was mainly due to a lack of consideration of the complex about data characteristics, overfitting, and incorrect management of missing data. Complexities of data should be handled appropriately. Competing risk can preclude the occurrence of the event of interest such as death before being diagnosed with liver cancer, which may reduce the risk of developing liver cancer [37]. Censored observation such as loss to follow-up in the prospective study may provide more information than we are already known. Simply ignoring the censored observations and analyzing the uncensored complete observations would lose efficiency and produce estimation bias [38]. Missing data is a common problem in the epidemiological studies. Most studies tended to exclude observations with missing value and use complete data or choose not to report it, which undoubtedly yielded bias and lost information. There are sophisticated methods to handle missing data, and the most common way is multiple imputation [39 – 41]. Only one study used multiple imputations methods to impute data with missing values [30], whereas many studies did not provide this information. Enough external and internal validation is of significance to increase the credibility of the model, which could increase the applicability of the model in a real-world practice and reduce the possibility of overfitting. However, only six studies and nine publications had external validation and internal validation, respectively. Failure of external validation restricted the general applicability undoubtedly.

Population selection of the derivation and validation process is very important to model applicability. A comprehensive and clear description of the study population could help to understand the observed variability among studies and provide information on model application. First, the source of the target population should be considered. The majority of studies in our systematic review were conducted on patients with hepatitis B surface antigen seropositive or with cirrhosis. Although HBV infection and cirrhosis are the common risk factors for liver cancer [42, 43], but the applicability of models based on these patients was limited. Two studies adopted a dataset from the average-risk population [29, 30], and one study was conducted in a population newly diagnosed with diabetes [21]. In Chi-Pang Wen et al, the prediction model could be used both in average-risk general public and in high-risk individuals, significantly increasing the application efficiency [29]. Sample size is the second main criteria when considering the representativeness of the entire population. Suitable sample size could precise the estimates [44]. EPV higher at least 20 are less likely to be overfitting. Unfortunately, only half of the studies could meet the criteria. A smaller sample size was observed in the external validation dataset.

Despite the variability of risk prediction models, a series of common predictors (including age, sex, cirrhosis, etc.) were incorporated into models, providing more information for future model development. Traditional risk factors of liver cancer combined with specific biomarkers could produce good discriminatory ability. Potential predictors were assessed which should be suitable for the targeted setting in which the model will be implemented. It is also necessary to note the accessibility and appropriateness of the possible predictors. Age and sex as common risk factors were nearly included for all risk prediction models. Objective predictors such as ALT, the ratio of aspirate aminotransferase to ALT, serum HCV RNA levels, and HCV genotype are susceptible to machine among different study sites, and doctor's experience. Core promoter mutations were not easily available in some centers, which might not be a general predictor when modeling [45]. Two ways to identify candidate predictors that should be taken into account are clinical knowledge and systematic reviews.

Model performance presented as discrimination and calibration is a key factor to assess the effectiveness of the model. Discrimination refers to how well the model differentiates between patients that will experience the outcome and who will not [46]. The discriminatory ability could be measured as AUC or C statistics. The highest discrimination in the included studies was 0.96 and the lowest discrimination was 0.61. These studies showed a good discrimination (0.7–0.8 is regarded as acceptable and 0.8–0.9 is excellent [47]. Calibration is an assessment of how closely the predictions of the model match the observed outcomes of the data [46], which was assessed using Hosmer-Lemeshow goodness of test usually. A poor calibration could result in overestimation (when the model predicts a higher risk than the actual observed risk) or underestimation (when the model predicts a lower risk than the actual observed risk) [48]. However, model calibration was poorly reported, which easily been questioned. PROBAST, a tool to assess the risk of bias and applicability of prediction model studies, which was published in 2019, recommended that model discrimination and calibration should be evaluated and reported sufficiently [16]. Further study is required to optimize the reporting of the prediction model measures based on this tool.

This review, which, to our knowledge, is the first study to synthesize model performance and to evaluate the methodological issues for liver cancer prediction models. The source of risk bias of the in-

cluded models was identified, which could provide basic information for model development and update in the future. In this study, a robust and reliable systematic methodology were used, which increased credibility. In addition, the current study was based on several international databases for analysis, which to some extent decreased the missing rate. However, several limitations of this study deserved attention. For instance, the strict inclusion and exclusion criteria were used. This may result in deleting studies with potential value. Further, confidence intervals were not acquired in some studies.

6. Conclusions

We identified 20 risk prediction models for liver cancer, and excellent discriminative ability was reported nearly in all models. However, these models in our systematic review were rated at high risk of bias. Therefore, there is greater room for the improvement on the risk prediction models for liver cancer, particular in the selection of representative population, data management and the reporting of model performance measures in the future.

7. Abbreviations

AUC: areas under curve; ALT: alanine aminotransferase; AST: aspirate aminotransferase; AFP: alpha-fetoprotein; BMI: body mass index; CHARM: the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies; CLD: chronic liver disease; EPV: every per events; FPG: fasting plasma glucose; GGT: gamma-glutamyl transpeptidase; HBV: hepatitis B virus; HBeAg: hepatitis B envelope antigen; HBsAg: hepatitis B surface antigen; PROBAST: the Prediction model Risk Of Bias Assessment Tool; PRIMA: the Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PPV: Positive predictive value; SVR: sustained virological response; TC: total cholesterol; VTQ: Virtual touch quantification; NPV: Negative predictive value; HL: Hosmer-Lemeshow.

8. Availability of Data and Materials

The data that support the findings of this study are extracted from existing publications which are also available on reasonable request from the corresponding author.

9. Funding

This work was funded by the Major State Basic Innovation Program of the Chinese Academy of Medical Sciences (2019-I2M-2-004), Sanming Project of Medicine in Shenzhen (SZSM201911015), National Natural Science Foundation of China (81974492).

10. Authors' Contributions

MM C, JZ Z, H L, DQ S, SY H, J L, N L, L L, J P and WQ C contributed to the study concept and design. MM C, H L screened titles and abstract for inclusion. MM C, H L, SY H, and DQ S contributed to the screening, data abstraction. H L, DQ S and YD Z contributed to quality assessment of included studies. MM C drafted the first manuscript. All authors contributed to the data interpretation and approved the final manuscript.

References:

1. lobal Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. 2020. https://gco.iarc.fr/today.
2. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. The Lancet Global Health. 2018; 6: 555-67.
3. Singal AG, Pillai A, Tiro J. Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. PLoS Med. 2014; 11: 1001624.
4. Kemp W, Pianko S, Nguyen S, Bailey MJ, Roberts SK. Survival in hepatocellular carcinoma: impact of screening and etiology of liver disease. J Gastroenterol Hepatol. 2005; 20: 873-81.
5. Tanaka H, Nouso K, Kobashi H, Kobayashi Y, Nakamura S, Miyake Y et al. Surveillance of hepatocellular carcinoma in patients with hepatitis C virus infection may improve patient survival. Liver Int. 2006; 26: 543-51.
6. McMahon BJ, Bulkow L, Harpster A, Snowball M, Lanier A, Sacco F et al. Screening for hepatocellular carcinoma in Alaska natives infected with chronic hepatitis B: a 16-year population-based study. Hepatology. 2000; 32: 842-6.
7. Noda I, Kitamoto M, Nakahara H, Hayashi R, Okimoto T, Monzen Y et al. Regular surveillance by imaging for early detection and better prognosis of hepatocellular carcinoma in patients infected with hepatitis C virus. J Gastroenterol. 2010; 45: 105-12.
8. Wong VW, Chan SL, Mo F, Chan TC, Loong HH, Wong GL et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. J Clin Oncol. 2010; 28: 1660-5.
9. Wong GL, Chan HL, Wong CK, Leung C, Chan CY, Ho PP et al. Liver stiffness-based optimization of hepatocellular carcinoma risk score in patients with chronic hepatitis B. J Hepatol. 2014; 60: 339-45.
10. Yang HI, Sherman M, Su J, Chen PJ, Liaw YF, Iloeje UH et al. Nomograms for risk of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. J Clin Oncol. 2010; 28: 2437-44.
11. Hung YC, Lin CL, Liu CJ, Hung H, Lin SM, Lee SD et al. Development of risk scoring system for stratifying population for hepatocellular carcinoma screening. Hepatology. 2015; 61: 1934-44.
12. Ensor J, Riley RD, Moore D, Snell KIE, Bayliss S, Fitzmaurice D et al. Systematic review of prognostic models for recurrent Venous Thromboembolism (VTE) post-treatment of first unprovoked VTE. BMJ Open. 2016; 6: 011190.
13. Gray EP, Teare MD, Stevens J, Archer R. Risk Prediction Models for Lung Cancer: A Systematic Review. Clin Lung Cancer. 2016; 17: 95-106.
14. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk Prediction Models for Colorectal Cancer: A Systematic Review. Cancer Prev Res. 2016; 9: 13-26.
15. Aladwani M, Lophatananon A, Ollier W, Muir K. Prediction models for prostate cancer to be used in the primary care setting: a systematic

review. BMJ Open. 2020; 10: 034661.

16. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019; 170: 51-8.
17. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014; 11: 1001744.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009; 6: 1000097.
19. Lee MH, Lu SN, Yuan Y, Yang HI, Jen CL, You SL et al. Development and validation of a clinical scoring system for predicting risk of HCC in asymptomatic individuals seropositive for anti-HCV antibodies. PloS One. 2014; 9: 94760.
20. Lee MH, Yang HI, Liu J, Batrla-Utermann R, Jen CL, Iloeje UH et al. Prediction models of long-term cirrhosis and hepatocellular carcinoma risk in chronic hepatitis B patients: risk scores integrating host and virus profiles. Hepatology. 2013; 58: 546-54.
21. Rau HH, Hsu CY, Lin YA, Atique S, Fuad A, Wei LM et al. Development of a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network. Comput Methods Programs Biomed. 2016; 125: 58-65.
22. Tseng TC, Liu CJ, Chen CL, Yang HC, Su TH, Wang CC et al. Risk stratification of hepatocellular carcinoma in hepatitis B virus e antigen-negative carriers by combining viral biomarkers. J Infect Dis. 2013; 208: 584-93.
23. Yang HI, Yuen MF, Chan HL, Han KH, Chen PJ, Kim DY et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. Lancet Oncol. 2011; 12: 568-74.
24. Yuen MF, Tanaka Y, Fong DY, Fung J, Wong DK, Yuen JC et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. J Hepatol. 2009; 50: 80-8.
25. Zhang H, Zhu J, Xi L, Xu C, Wu A. Validation of the Toronto hepatocellular carcinoma risk index for patients with cirrhosis in China: a retrospective cohort study. World J Surg Oncol. 2019; 17: 75.
26. El-Serag HB, Kanwal F, Davila JA, Kramer J, Richardson P. A new laboratory-based algorithm to predict development of hepatocellular carcinoma in patients with hepatitis C and cirrhosis. Gastroenterology. 2014; 146: 1249-55.
27. Ioannou GN, Green P, Kerr KF, Berry K. Models estimating risk of hepatocellular carcinoma in patients with alcohol or NAFLD-related cirrhosis for risk stratification. J Hepatol. 2019; 71: 523-33.
28. Singal AG, Mukherjee A, Elmunzer BJ, Higgins PD, Lok AS, Zhu J et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol. 2013; 108: 1723-30.
29. Wen CP, Lin J, Yang YC, Tsai MK, Tsao CK, Etzel C et al. Hepatocellular carcinoma risk prediction model for the general population: the predictive power of transaminases. J Natl Cancer Inst. 2012; 104: 1599-611.
30. Chung JW, Jang ES, Kim J, Jeong SH, Kim N, Lee DH, et al. Development of a nomogram for screening of hepatitis B virus-associated hepatocellular carcinoma. Oncotarget. 2017; 8: 106499-510.
31. Sinn DH, Kang D, Cho SJ, Paik SW, Guallar E, Cho J et al. Risk of hepatocellular carcinoma in individuals without traditional risk factors: development and validation of a novel risk score. Int J Epidemiol. 2020; 49: 1562-71.
32. Aoki T, Iijima H, Tada T, Kumada T, Nishimura T, Nakano C 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: 104-12.
33. Michikawa T, Inoue M, Sawada N, Iwasaki M, Tanaka Y, Shimazu T et al. Development of a prediction model for 10-year risk of hepatocellular carcinoma in middle-aged Japanese: The Japan Public Health Center-based Prospective Study Cohort II. Prev Med. 2012; 55: 137-43.
34. Flemming JA, Yang JD, Vittinghoff E, Kim WR, Terrault NA. Risk prediction of hepatocellular carcinoma in patients with cirrhosis: the ADRESS-HCC risk model. Cancer. 2014; 120: 3485-93.
35. Ganne-Carrie N, Layese R, Bourcier V, Cagnot C, Marcellin P, Guyader D et al. Nomogram for individualized prediction of hepatocellular carcinoma occurrence in hepatitis C virus cirrhosis (ANRS CO12 CirVir). Hepatology. 2016; 64: 1136-47.
36. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg. 2010; 126: 2234-42.
37. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009; 170: 244-56.
38. Leung KM, Elashoff RM, Afifi AA. Censoring issues in survival analysis. Annu Rev Public Health. 1997; 18: 83-104.
39. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol. 2006; 59: 1087-91.
40. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017; 9: 157-66.
41. de Goeij MC, van Diepen M, Jager KJ, Tripepi G, Zoccali C, Dekker FW. Multiple imputation: dealing with missing data. Nephrol Dial Transplant. 2013; 28: 2415-20.
42. Mak LY, Cruz-Ramon V, Chinchilla-Lopez P, Torres HA, LoConte NK, Rice JP et al. Global Epidemiology, Prevention, and Management of Hepatocellular Carcinoma. Am Soc Clin Oncol Edu Book. 2018; 38: 262-79.
43. Iavarone M, Colombo M. HBV infection and hepatocellular carcinoma. Clin Liver Dis. 2013; 17: 375-97.
44. Weber EJ, Hoo ZH. Why sample size estimates? Emerg Med J. 2018; 35: 755.
45. Yuen M-F, Tanaka Y, Fong DY-T, Fung J, Wong DK-H, Yuen JC-H et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. J Hepatol. 2009; 50: 80-8.

46. Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model†. Eur J Cardiothorac Surg. 2018; 54: 203-8.

47. Brown FS, Glasmacher SA, Kearns PKA, MacDougall N, Hunt D, Connick P et al. Systematic review of prediction models in relapsing remitting multiple sclerosis. PloS One. 2020; 15: e0233575.

48. Carrillo-Larco RM, Aparcana-Granda DJ, Mejia JR, Barengo NC, Bernabe-Ortiz A. Risk scores for type 2 diabetes mellitus in Latin America: a systematic review of population-based studies. Diabet Med. 2019; 36: 1573-84.

49. Demirtas CO, Gunduz F, Kani HT, Keklikkiran C, Alahdab YO, Yilmaz Y et al. External validation of the Toronto hepatocellular carcinoma risk index in Turkish cirrhotic patients. Eur J Gastroenterol Hepatol. 2020; 32: 882-8.