REVIEW

# Systematic review: risk prediction models for recurrence of hepatocellular carcinoma after liver transplantation

Abdulahad Abdulrab Mohammed Al-Ameri[1,2,3], Xuyong Wei[1,2,3], Xue Wen[1,2,3], Qiang Wei[1,2,3], Haijun Guo[1,2,3], Shusen Zheng[1,2,3] & Xiao Xu[1,2,3]

1 Department of Hepatobiliary and Pancreatic Surgery, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
2 Institution of Organ Transplantation, Zhejiang University, Hangzhou, China
3 NHFPC Key Laboratory of Combined Multi-organ Transplantation, Hangzhou, Zhejiang Province, China

**Correspondence**
Xiao Xu MD, PhD, Department of Hepatobiliary and Pancreatic Surgery, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China.
Tel: +86-571-87236567;
fax: +86-571-87236567;
e-mail: zjxu@zju.edu.cn

## SUMMARY

Recurrence of hepatocellular carcinoma (HCC) after liver transplantation (LT) is a significant clinical problem associated with poor surgical outcomes. This study aims to summarize the current evidence on risk prediction models of HCC recurrence after LT. PubMed and EMBASE were searched to May 25, 2019, for relevant articles. Studies originally designed to develop or validate a risk prediction model for HCC recurrence after LT were included. Two independent authors summarized the study characteristics and evaluated the risk of bias and applicability concerns in the included studies. From 26 included studies, 18 original risk prediction models were determined, but only five models were externally validated. The average number of predictors involved in the construction of risk models was three. The most frequently employed predictors were alpha-fetoprotein, tumor size, vascular invasion, tumor number, tumor differentiation, and neutrophil-lymphocyte ratio. Most studies showed good discriminatory performance (AUC >0.75). The overall quality of the included studies was generally low. Most of the original models lacked the highly recommended external and prospective validation in diverse populations. The AFP model was the well-validated preoperative risk model that can stratify patients into high- and low-risk groups.

*Transplant International* 2020; 33: 697–712

**Key words**
hepatocellular carcinoma, liver transplantation, prediction, prognosis, recurrence, risk prediction model

Received: 5 September 2019; Revision requested: 10 October 2019; Accepted: 21 January 2020; Published online: 25 February 2020

## Introduction

Hepatocellular carcinoma (HCC), the most common primary liver cancer, is a significant health challenge for doctors and patients worldwide. According to the World Health Organization, more than one million patients are expected to die because of liver cancer within the next decade [1]. Liver transplantation (LT) is the best therapeutic modality in a group of HCC patients with underlying cirrhosis. However, the recurrence of HCC is a significant clinical problem for 8–30% of all patients who have undergone LT [2,3].

The development of post-LT HCC recurrence appears to be multifactorial, with numerous pre-, peri-, and post-transplant predictors. Several studies have attempted to outline the role of these various predictors


doi:10.1111/tri.13585

to develop risk prediction models for HCC recurrence after LT. The risk prediction models have potential applicability for individual-based stratification and can be used to develop therapeutic algorithms, to guide treatment decisions, and to advise patients. This study provides the first comprehensive review of risk scoring systems originally designed to predict HCC recurrence in patients who had undergone LT.

## Materials and methods

This study was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [4], the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) [5], and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [6] where they were applicable. The PRISMA checklist is available in the Supporting Information (Table S1). Patient informed consent and ethical approval were not required, because all data were extracted from online articles. The systematic literature review was performed according to a priori established study protocol.

### Search strategy

Two independent researchers (AA and XW) performed an electronic bibliographic search of PubMed and EMBASE without time restrictions, using a combination of the following keywords: HCC, hepatocellular carcinoma, liver cancer, hepatoma, recurrence, relapse, model, score, nomogram, risk, prediction, prognosis, and liver transplantation. The complete search strategies are presented in the Supporting Information (Data S1). Only studies of humans and studies published in English were considered for inclusion. We then scanned references from the links of related studies in PubMed to identify articles that fulfilled the inclusion criteria. To identify additional relevant studies, we performed another manual screening for the bibliographies of the involved studies and related reviews. Studies that were published as abstracts were excluded from the analysis. The last search date was May 25, 2019.

### Inclusion criteria

The studies were included on the basis of the following predefined criteria: (i) Original research article published in a peer-reviewed journal; (ii) only articles with participants who had undergone LT; (iii) studies that confirmed that HCC was pathologically diagnosed after LT; (iv) studies that identified risk factors for developing post-LT HCC recurrence and constructed or validated a risk prediction model at the individual level; (v) studies that provided a risk measure of HCC recurrence using a combination of two or more risk factors to predict the risk of post-LT recurrence; and (vi) studies that defined the recurrence of HCC after LT as a primary outcome. Studies were excluded if they were originally designed to predict overall survival rather than recurrence or if they were in the form of a conference report or abstract.

### Data extraction and quality assessment

After the inclusion of the studies, data were extracted independently by two authors (AA and XW). The obtained data included the first author name, publication year, study region, study period, study type (including study design and whether the study was conducted in multiple centers or a single center), number of patients, study outcome, area under curve (AUC)/*C*-statistic or net reclassification index (NRI), calibration, accuracy measures, standard selection criteria, and the underlying etiology of the liver disease. We then classified studies into the following types according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines[6]: development only (1a), development and validation using resampling (1b), random (2a) or nonrandom (2b) split-sample development and validation, development and validation using separate data (3), or validation only (4). The level of evidence was estimated according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria [7]. The risk of bias (ROB) and applicability concerns of development and validation studies were estimated based on the Prediction Study Risk of Bias Assessment Tool (PROBAST) [8], a novel assessment tool for ROB, which had been precisely designed for systematic reviews of prediction studies. This newly launched tool has three separate domains: participant selection, predictors, and outcome, with 23 signaling questions that classify ROB into high, low, or unclear. It also assesses the applicability of a model. Other extracted data included the presentation method of the final risk prediction model, the model type (preoperative, postoperative, or general), assessment methods of involved risk factors, risk factors involved in the final model, and risk factors



Transplant International 2020; 33: 697–712

considered by the univariate analysis but not included in the final model. In case of any discrepancy, an agreement was obtained via discussion with senior authors (ZS, XX).

### Data synthesis

Considering a large number of prediction models and heterogeneity in characteristics of the included studies, we selected a narrative synthesis of results supported by tables, with characteristics listed for every included study. The discriminatory performance of a risk prediction model for HCC recurrence after LT was measured using the AUC/C statistics or NRI. This indicator shows how well the model differentiates between patients with high and low risk of HCC recurrence after LT. AUCs, NRI, calibration and accuracy measures (sensitivity, specificity, negative and positive predictive values) of development and validation cohorts were reported if the data were available in the published papers. AUCs range from 0.5 (poor discrimination) to 1 (perfect discrimination) [9,10]. Based on the population characteristics of the validation cohort, the *C*-statistic of the same prediction model can vary greatly, especially because of the heterogeneity of the population from which the predictors of the risk prediction model were derived [11]. An AUC of >0.75 is considered an indicator of good discriminatory performance [12]. The NRI is a statistical measure used to evaluate whether a new risk model can provide clinically relevant improvements in the discriminatory performance of an old model [13].

## Results

### Study selection

Our search strategy identified a total of 257 records. After the removal of duplicates ($n$ = 105) and exclusion at the title and abstract level ($n$ = 71), 81 papers were qualified for full-text evaluation. After full-text review, 55 articles were excluded for the following reasons: The primary outcome was not HCC recurrence after LT ($n$ = 47); the risk score was unavailable (n = 6); or the model only predicted survival after HCC recurrence

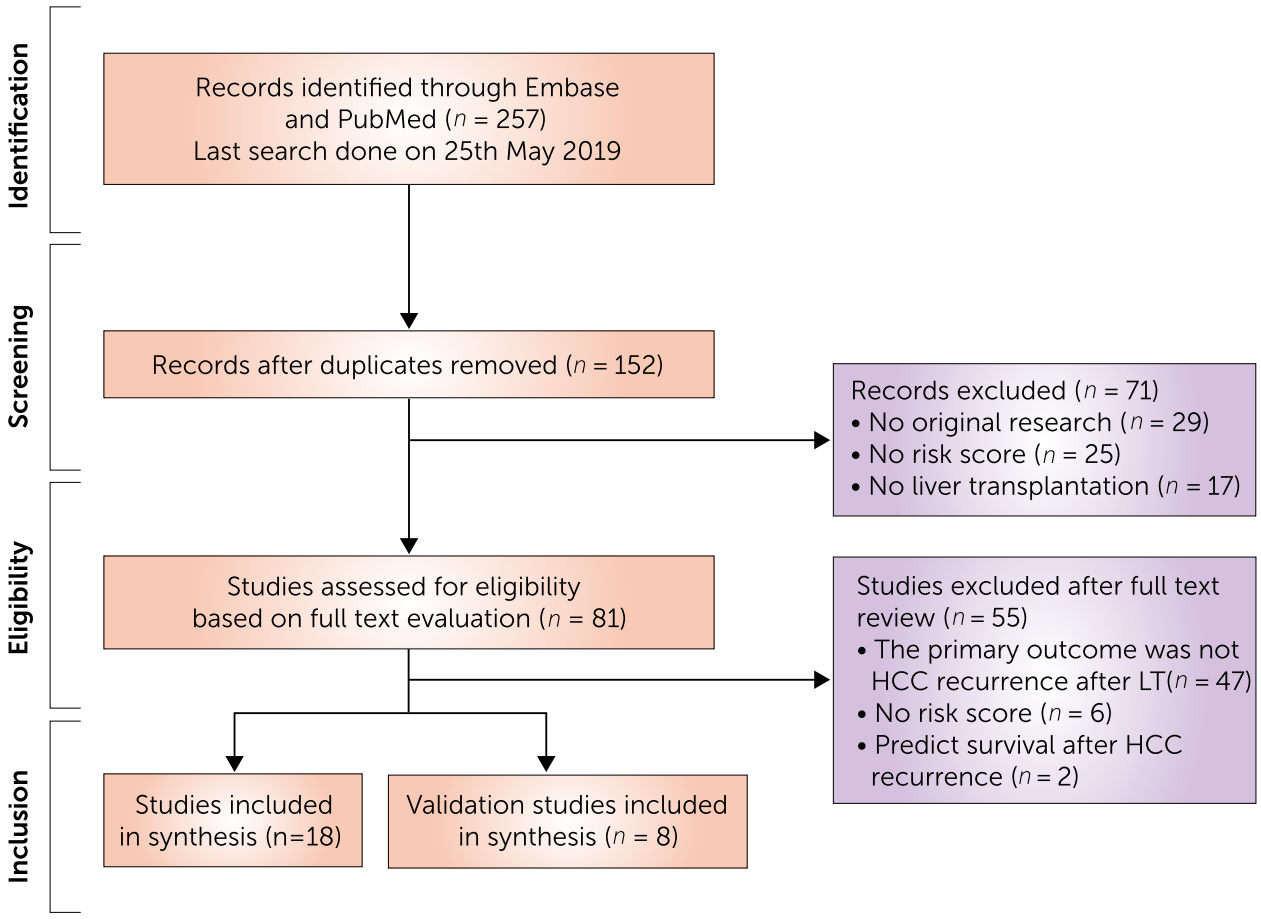


**Figure 1** Preferred reporting items for systemic reviews and meta-analysis (PRISMA) flowchart.

($n = 2$). This left 26 studies [14–39] for inclusion in the final analysis, including 18 original risk prediction models and eight validation studies. The details of the selection procedure are described in Fig. 1.

### Characteristics of development and validation studies

The characteristics of the 18 original risk prediction models and their validation studies are presented in Tables 1 and 2. Only seven studies came from hepatitis B virus (HBV) epidemic regions, while the remaining studies ($n = 11$) originated in western countries, where hepatitis C virus (HCV) is the most common cause of HCC. Only one of the 18 studies reported on younger patients [31]. The study periods had very long time intervals, ranging from 1981 to 2016. The population size greatly differed among studies and ranged from 75 to 3276 patients. Fifteen of the 18 original risk models reported a 5-year recurrence rate as the primary outcome. Two studies reported a three-year recurrence rate [19,25], and one study presented a two-year recurrence rate [16].

The AUC/C statistics were described in most studies except nine [15,25,30–36]. The NRI was described in six studies [20,26,33–35,39] to quantify how well a new model can reclassify patients with HCC recurrence compared with Milan criteria; however, the 95% CI was not provided except in Rhu et al.'s study [35]. Most models showed good discriminatory performance with AUC/C statistics >0.75 either in the development or validation cohorts. Only in nine risk prediction models were the AUC/C statistics < 0.75 either in development or validation cohorts, indicating modest discriminatory performance [14,17,20–22,26,28,29,31]. Of the 18 original risk prediction models, calibration was reported in only five studies [14,16,21,23,26] and was presented as a curve in two studies, as a Hosmer–Lemeshow statistic in another two studies, or as a chi-squared test in only one study. Sensitivity and specificity were reported in six risk prediction models and ranged from high sensitivity (100%) and low specificity (7%) in Chan et al.'s model [29] to low sensitivity (26.1%) and high specificity (91.8%) in Lai et al.'s model [22].

The presentation and assessment methods of the risk factors involved in the final models and the risk factors that were considered by the univariate analysis but were not involved in the final risk scores are described in Table 3. Half the original risk models were presented in a point-based format without any available online calculation tools. The majority of the models were constructed based on traditional risk factors of HCC recurrence, including alpha-fetoprotein (AFP) [14–22,24], tumor size [15,16,18,24,26,28–30], vascular invasion [18,20,24,27,29–31], tumor number [15,18,26–28], and tumor differentiation [16,24,29,30]. Other risk models were derived using a combination of these well-known predictors with other serological biomarkers, such as the neutrophil-lymphocyte ratio (NLR) [18,22,24,27], fibrinogen levels [17,25], and cholesterol [24]. The average number of predictors involved in the construction of risk models was three.

Only five of the 18 original risk scores were validated [20,26,28,29,31]. The AFP model [26] was the most validated risk model in five studies, followed by Decan et al.'s model [28] in four studies, Mehta et al.'s model [20] in two studies, and Iwatsuki et al.'s model [31] in two studies. No considerable difference was observed in the NRI of the AFP model of Duvoux et al. [26], which ranged between 0.02 in Rhu et al.'s study [35] from Korea and 0.06 in the study of Piñero et al. [33] from Latin America. The other characteristics of the validation studies are illustrated in Table 2.

### Quality assessment

Among the 26 included studies, 25 were retrospective cohort studies except for Halazun et al.'s prospective study [18]. The level of evidence according to GRADE criteria showed that all studies of the included original risk models were classified as low quality (Table 1). The PROBAST tool was used to evaluate the ROB and applicability concerns in the involved risk prediction models and their validation studies (Table 4). Overall, only eight of the 26 studies were considered with low ROB and applicability concerns [18,20,22,24,26,28,34,39]. For the patient selection domain, 18 of the 26 studies clearly defined the inclusion and exclusion criteria for study participants [16–23,25–28,31,33–35,37,39]. The remaining studies either did not offer detailed data for this domain, or they were unclear or had a smaller sample size; thus, the ROB and applicability concerns were reported as high or unclear. The determination of outcome without a knowledge of predictor information was not clearly reported in 12 of the 18 the original risk models; thus, this domain was assessed as unclear. For the analysis domain, the majority of the studies had no missing data except the study of Ling et al. [16], which was handled appropriately. However, the analysis domain was reported as low in 12 of the 26 included studies, either because the predictor selection was based on univariable analysis and/or the complexities of the

**Table 1.** Characteristics of the original risk prediction models.

| Author, year | Study region | Study period | Study type | No. of patients | Outcomes | NRI† or AUC/C statistics (95%CI) | Calibration | Accuracy (%) | Standard selection criteria | Etiology (%) | TRIPOD level* | GRADE score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ma, 2019 [14] | China | 1995–2016 | Observational retrospective Single center | 183(D) 147(V) | 5-Y RFS, OS | 0.67 (0.55–0.79) 0.85 (0.74–0.96) | $\chi^2$, $P$ = 0.459 | _ | Milan, UCSF | HBV (81.6) Others (18.4) | 2a | Low |
| Shimamura, 2019 [15] (5-5-500 rule) | Japan | 1998–2009 | Observational retrospective Multicenter | 965 | 5-Y RR, OS | NA | _ | _ | Milan | HCV (60.3) HBV (29.2) Others (10.5) | 1a | Low |
| Ling, 2018 [16] | China | 2010–2015 | Observational retrospective Multicenter | 673(D) 337(V) | 1,2 Y RR And CR | 0.77 (0.74–0.80) 0.77 (0.73–0.80) 0.76 (0.69–0.83) (D) 0.86 (0.82–0.90) 0.86 (0.82–0.90) 0.84 (0.75–0.93) (V) | Curve | _ | Milan, Hangzhou | HBV (100?) | 2a | Low |
| Jiang, 2018 [17] | China | 2003–2009 | Observational retrospective Multicenter | 119(D) 109(V) | 5-Y RFS, OS | 0.82 (0.73–0.91) (D) 0.70 (0.60–0.80) (V) | _ | SS 77.0 SP 62.5(V) | Milan, Hangzhou | HBV (100?) | 3 | Low |
| Halazun, 2017 [18] (MORAL score USA) | USA | 2001–2012 | Observational prospective Single center | 339 | 5-Y RFS | 0.82 (0.77–0.88) (pre) 0.88 (0.83–0.93) (post) 0.91 (0.87–0.95) (general) | _ | _ | Milan, UCSF | HCV (69.3) HBV (15.3) Alcohol (11.8) Others (3.6) | 1a | Low |
| Feng, 2017 [19] | China | 2009–2013(D) 2005–2008(V) | Observational retrospective Single center | 117(D) 51(V) | 0.5,1,2, 3-Y RFS | 0.77 (0.65–0.89) (D) 0.84 (0.72–0.97) (V) | _ | _ | Milan? | HBV (100) | 3 | Low |
| Mehta, 2016 [20] (RETREAT score) | USA | 2002–2012 | Observational retrospective Multicenter | 721(D) 340(V) | 1,2,5-Y RR, OS | 0.77 (0.71–0.82) (D) 0.82 (0.77–0.86) (V) 1-Y† 0.40 5-Y† 0.31 | _ | _ | Milan | HCV (58) HBV (18.3) Alcohol (10.1) NASH (7.0) | 3 | Low |
| Piñero, 2016 [21] | Argentina | 2005–2011 | Observational retrospective Multicenter | 87 | 5-Y RR, OS | 0.74 (0.58–0.88) | H-L, $P$ = 0.96 | SS 62 SP 82 PPV 0.43 NPV 0.90 | Milan, UCSF, up-to-7 | HCV (41.1) HBV (11.6) Alcohol (19.7) Others (27.6) | 1b | Low |
| Lai 2016 [22] (TRAIN score) | Italy Belgium | 2000–2012(D) 2005–2014(V) | Retrospective Multicenter | 179(D) 110(V) | 5-Y RR, ITT-S | 0.58 (0.37–0.79) (D) 0.59 (0.37–0.81) (V) | _ | SS 26.1 SP 91.8 | Milan | HCV (47.1) HBV (11.6) Others (41.3) | 3 | Low |
| Lee, 2016 [23] (MORAL score Korea) | Korea | 2001–2013(D) 2003–2013(V) | Observational retrospective Multicenter Medical records | Out milan 92(D) 113(V) In milan = 361 | 1,5-Y RR, OS | 0.83 (0.73–0.92) (D) 0.80 (0.72–0.88) (V) | Curve | _ | Milan | HBV (87.8) HCV (6.9) Alcohol (2.1) Others (3.2) | 1b | Low |
| Agopian, 2015 [24] | USA | 1984–2013 | Observational retrospective Multicenter database | 865 | 1,3,5-Y RFS, OS | 0.79 (0.75–0.83) (pre) 0.85 (0.82–0.89) (post) | _ | _ | Milan, UCSF | HCV (58) HBV (16) Alcohol (9) Others (71) | 3 | Low |

**Table 1.** Continued.

| Author, year | Study region | Study period | Study type | No. of patients | Outcomes | NRI† or AUC/C statistics (95%CI) | Calibration | Accuracy (%) | Standard selection criteria | Etiology (%) | TRIPOD level* | GRADE score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Li, 2015 [25] | China | 2003–2010 | Observational retrospective Multicenter | 102 | 1,3-Y RFS | 0.66 (0.60–0.71) | _ | SS 55.3 SP 70.3 | Milan | HBV (87.8) Others (12.2) | 1a | Low |
| Duvoux, 2012 [26] (AFP model) | France | 1988–2001(D) 2003–2004(V) | Observational retrospective Multicenter databank | 537(D) 435(V) | 5-Y RR, OS | 0.70 (0.63–0.76) 0.11† | H-L, $P = 0.78$ | SS 53.4 SP 81.6 | Milan | Virus (50.9) Alcohol (35.6) Others (13.5) | 3 | Low |
| Wang, 2011 [27] | China | 2003–2009 | Observational retrospective Single center | 101 | 5-Y RFS, OS | 0.78 (0.69–0.88) | _ | _ | Milan, UCSF, Hangzhou | HBV (100) | 1a | Low |
| Decaens, 2010 [28] | France | 1988–1998(D) 1999–2001(V) | Observational retrospective Single center | 328(D) 140(V) | 5-Y RFS | 0.65 (0.59–0.71) | _ | | Milan | Virus (57.7) Alcohol (31.6) Others (10.7) | 3 | Low |
| Chan, 2008 [29] | USA | 2001–2005(D) 1996–2000(V1) 2002–2005(V2) | Observational retrospective Multicenter | 116(D) 31(V1) 41(V2) | 5-Y RR | 0.91(D) 1.00 (V1) 0.95 (V2) | _ | SS 100 SP 27 PPV 30 NPV 70 (D), SS 100 SP 7 PPV 60 NPV 40 (V1), SS 75 SP 11 PPV 43 NPV 57 (V2) | NA | HCV (62.1) HBV (18.1) Alcohol (16.4) Others (3.4) (D) | 3 | Low |
| Parfitt, 2007 [30] | UK | 1985–2003 | Observational retrospective Single center | 75 | 3,5,8-Y RFS, RR | | _ | SS = 55%, SP = 98% | Milan, UCSF | HCV (42.7) HBV (17.3) Alcohol (21.3) Others (18.7) | 1a | Low |
| Iwatsuki, 2000 [31] | USA | 1981–1998 | Observational retrospective Single center | 344 | 5-Y RFS | NA | _ | | NA | HCV (47.5) HBV (21.8) Others (30.7) | 1a | Low |

AUC, area under the curve; CI, confidence interval; CR, cumulative recurrence; D, development cohort; HBV, hepatitis B virus; HCV, hepatitis C virus; H-L, Hosmer Lemeshow C test; ITT-S, intention-to-treat survival; NA, not available; NPV, negative predictive value; NRI, net reclassification index; PPV, positive predictive value; pre, preoperative, post, postoperative; RFS, recurrence-free survival; RR, recurrence rate; SP, specificity; SS, sensitivity; UCSF, University of California, San Francisco; V, validation cohort.

†NRI.

*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, nonrandom split-sample development and validation; 3, development and validation using separate data; and 4, validation study.

**Table 2.** Characteristics of validation studies of original risk prediction models.

| Original model study | Validation study | Study region | Study period | Study type | No. of patients | Etiology (%) | Outcomes | NRI† or AUC/C statistics (95%CI) |
|---|---|---|---|---|---|---|---|---|
| Duvoux 2012 [26] | Duvoux 2012 [26] | France | 2003–2004 | Observational retrospective Multicenter | 435 | Virus (50.9) Alcohol (35.6) Others (13.5) | 5-Y RR, OS | 0.15† |
| | Varona 2015 [32] | Spain | 1996–2012 | Observational retrospective Single center | 109 | HCV (48) Alcohol (45) Others (7) | 1, 3, 5, 10, and 15-Y survival | NA† |
| | Piñero 2016 [33] | Latin America | 2005–2011 | Observational retrospective Multicenter | 327 | HBV (28.7) HCV (27.2) Alcohol (17.7) Others (26.4) | 1,3,5-Y RR, OS | 0.06† |
| | Notarpaolo 2017 [34] | Italy | 2002–2010 | Observational retrospective Multicenter | 574 | HCV (58.7) HBV (24) Alcohol (11.7) Others (5.6) | 5-Y RR, OS | 0.03† |
| | Rhu 2018 [35] | Korea | 2007–2015 | Observational retrospective Single center | 400 | HBV (83.8) HCV (10.2) Alcohol (4) Others (2) | 5-Y RR, OS | 0.02† (0.01–0.02) |
| Decaens 2010 [28] | Decaens 2010 [28] | France | 1988–1998(D) 1999–2001(V) | Observational retrospective Multicenter | 328(D) 140(V) | Virus (57.7) Alcohol (31.6) Others (10.7) | 5-Y RR, RFS | 0.65 (0.59–0.71) |
| | Marelli 2008 [36] | UK | 1988–2006 | Observational retrospective Single center | 100 | HCV (44) HBV (24) Alcohol (15) Others (17) | 3 months, 1,3,5-Y RFS, OS | NA |
| | Costentin 2017 [37] | France | 2003–2005 | Observational retrospective Multicenter | 321 | Virus (61.8) Alcohol (28.8) Others (9.4) | 5-Y RR | 0.75 (0.69–0.81) |
| | Fernández 2018 [38] | Spain | 2006–2015 | Observational retrospective Single center | 105 | HCV (39) HBV (16.2) Alcohol (15.2) Others (2) | 5-Y RR, RFS | 0.67 (0.48–0.87) |
| Chan 2008 [29] | Chan 2008 [29] | USA | 1996–2000(V1) 2002–2005(V2) | Observational retrospective Multicenter | 31(V1) 41(V2) | NA | 5-Y RR | 1.00 (0.0–0.98) (V1) 0.95 (0.89–0.98) (V2) |
| | Costentin 2017 [37] | France | 2003–2005 | Observational retrospective Multicenter | 321 | Virus (61.8) Alcohol (28.8) Others (9.4) | 5-Y RR | 0.67 (0.61–0.97) |
| | Fernández 2018 [38] | Spain | 2006–2015 | Observational retrospective Single center | 105 | HCV (39) HBV (16.2) Alcohol (15.2) Others (2) | 5-Y RR, RFS | 0.81(0.65-0.98) |

**Table 2.** Continued.

| Original model study | Validation study | Study region | Study period | Study type | No. of patients | Etiology (%) | Outcomes | NRI† or AUC/C statistics (95%CI) |
|---|---|---|---|---|---|---|---|---|
| Mehta 2016 [20] | Mehta 2016 [20] | USA | 2002–2012 | Observational retrospective Multicenter | 340 | HCV (58)<br>HBV (18.3)<br>Alcohol (10.1)<br>Others (13.6) | 1,2,5-Y RR, OS | 0.82 (0.77–0.86) |
| | Mehta 2018 [39] | USA | 2012–2014 | Observational retrospective Multicenter | 3276 | HCV (62.8)<br>HBV (5.5)<br>Alcohol (7.3)<br>Others (24.4) | 1,3-Y RR, OS | 0.75 (0.71–0.79)<br>1-Y† 0.17<br>3-Y† 0.28 |
| Iwatsuki 2000 [31] | Iwatsuki 2000 [31] | USA | 1981–1998 | Observational retrospective Single center | 344 | HCV (47.5)<br>HBV (21.8)<br>Others (30.7) | 5-Y RFS | NA |
| | Costentin 2017 [37] | France | 2003–2005 | Observational retrospective Multicenter | 365 | Virus (61.8)<br>Alcohol (28.8)<br>Others (9.4) | 5-Y RR | 0.71 (0.65–0.77) |

AUC, area under the curve; CI, confidence interval; D, development cohort; HBV, hepatitis B virus; HCV, hepatitis C virus; NA, not available; NRI, net reclassification index; RFS, recurrence-free survival; RR, recurrence rate; V, validation cohort.
†NRI.

data were not sufficiently clarified, except in 14 studies [15,18,20,22–24,26,28,29,33–35,37,39].

## Discussion

Since the revolution of the Milan criteria in 1996, it has been considered a benchmark for the selection of HCC candidates for LT [40–42]. However, the limitations on the tumor metrics render the Milan criteria very restrictive, and some HCC candidates with a potentially good prognosis lose the opportunity to benefit from LT. Thus, new criteria (e.g., UCSF [43], Hangzhou [44], and up-to-seven criteria [45]) have been proposed and validated to expand the pool of HCC candidates with a similar prognosis to that of the Milan criteria [46]. However, this expansion usually occurs at the expense of HCC recurrence. So several risk models have been developed to predict HCC recurrence in patients who have undergone LT.

To the best of our knowledge, this study is the first comprehensive systematic review of risk prediction models for HCC recurrence after LT. A total of 26 studies, including 18 original risk prediction models, were evaluated. The construction of models depends on applying multivariate analysis if the variables have significant results according to the univariate analysis ($P < 0.05$). The conversion of continuous variables into categorical variables was observed in many studies, which may have resulted in information loss and reduced statistical power to detect an association between the risk factors and study outcome. Thus, converting continuous variables into categorical variables should not be performed except for convincing reasons, and if the conversion is inevitable, then the cutoff values should be applied.

To make clear recommendations for the selection of the most appropriate risk prediction model, we should take into account the study design, overall quality, simplicity, and external validation and how well a predicted probability of HCC recurrence matches the observed risk at the individual level (Table 5). The estimation of risk prediction models should not be limited only to ROC analysis, but other measures (e.g., calibration and accuracy) should also be evaluated. In our analysis, although eight risk models [14,16–20,24,29] showed a good discriminatory performance (AUC >0.80), the calibration was observed only in five risk models. The results also showed that 13 of the original risk models [14–19,21–25,27,30] lacked either internal or external validation, so they cannot be recommended in clinical practice even if they have a good quality and good

**Table 3.** Model presentation, type, assessment, and risk factors that were included or considered in the construction of predictive models.

| | | | | Risk factors | | |
|---|---|---|---|---|---|---|
| Author, year | Model presentation | Model type | Assessment | Total no. | Factors included in the final model | Factors considered by univariate analysis but not included in the final model |
| Ma, 2019 [14] | Equation | General | Pathology<br>Serology | 3 | Serum AFP, sum of largest tumor size and number and salvage LT | Age, platelet count, graft (g)/ESLW, Milan, UCSF criteria |
| Shimamura, 2019 [15] (5-5-500 rule) | Equation | Preoperative | Radiology<br>Serology | 3 | AFP, tumor size and nodule number | NA |
| Ling, 2018 [16] | Nomogram | General | Pathology<br>Serology | 4 | AFP, differentiation, total tumor size and CIT | Tumor number, corticosteroid usage/donor: age, anti-Hbc, DWIT, anhepatic time, ABO incompatible |
| Jiang, 2018 [17] | Equation | Preoperative | Serology | 2 | Serum AFP and fibrinogen | Age, TTV |
| Halazun, 2017 [18] (MORAL score USA) | Point-based model | Preoperative<br>Postoperative<br>General | Pathology<br>Radiology<br>Serology | 3<br>3<br>6 | AFP, NLR, largest tumor size) AND/ OR (vascular invasion, largest tumor size, tumor number | TTV, tumor no. on radiology, tumor necrosis |
| Feng,2017 [19] | Point-based model | Postoperative | Pathology | 3 | AFP, CK19, GPC3 | MVI, histological grading |
| Mehta, 2016 [20] (RETREAT score) | Point-based model | General | Pathology<br>Serology | 3 | Serum AFP, sum of largest tumor size and number and MVI | Tumor differentiation, wait time |
| Piñero, 2016 [21] | Point-based model | Preoperative | Radiology<br>Serology | 2 | Serum AFP and Up-to-7 imaging criteria | Waiting list, UCSF, Milan, MVI, tumor differentiation |
| Lai, 2016 [22] (TRAIN score) | Equation | General | Pathology<br>Radiology<br>Serology | 4 | AFP slope, mRECIST-PD, NLR and WT | NA |
| Lee, 2016 [23] (MORAL score Korea) | Equation | Preoperative | serology | 2 | AFP, PIVKA-II | Largest tumor size, tumor number, tumor type |

**Table 3.** Continued.

| Author, year | Model presentation | Model type | Assessment | Risk factors | | |
|---|---|---|---|---|---|---|
| | | | | Total no. | Factors included in the final model | Factors considered by univariate analysis but not included in the final model |
| Agopian, 2015 [24] | Nomogram | General | Pathology<br>Radiology<br>Serology | 7 | Tumor differentiation, vascular invasion, downstaging, largest tumor diameter. AFP, NLR, and total cholesterol | Age >55-year, hepatitis C, meld >10 |
| Li, 2015 [25] | Equation | Preoperative | Radiology<br>Serology | 3 | HBsAg, plasma fibrinogen and TTV | Preoperative NLR, vascular invasion |
| Duvoux, 2012 [26] (AFP model) | Point-based model | Preoperative | Radiology<br>Serology | 3 | Pre-LT AFP, largest tumor size and tumor number | Underlying cirrhosis, liver disease, preoperative HCC treatment, Milan criteria, tumor location, gamma-glutamyltransferase level, alkaline phosphatase level |
| Wang, 2011 [27] | Point-based model | Preoperative | Radiology<br>Serology | 3 | Macrovascular invasion, tumor number, NLR | Largest tumor size, AFP, HBV-DNA level |
| Decaens, 2010 [28] | Point-based model | Preoperative | Radiology<br>Serology | 3 | Largest tumor diameter, number of nodules and tumor differentiation | Underlying cirrhosis and liver disease, gamma-glutamyltransferase level, AFP, alkaline phosphatase, pre- or post-LT HCC treatment, tumor volume, tumor location |

**Table 3.** Continued.

| Author, year | Model presentation | Model type | Assessment | Risk factors | | |
|---|---|---|---|---|---|---|
| | | | | Total no. | Factors included in the final model | Factors considered by univariate analysis but not included in the final model |
| Chan, 2008 [29] | Point-based model | Postoperative | Pathology | 4 | Largest tumor diameter, bilobar extension, macrovascular invasion, tumor differentiation | multilobulated HCC, total number of tumors, total tumor size |
| Parfitt, 2007 [30] | Nomogram | Postoperative | Pathology | 5 | MVI, tumor size, tumor differentiation, microsatellitosis, and giant/bizarre cells | AFP, cryptogenic cirrhosis, OKT3, tumor necrosis |
| Iwatsuki, 2000 [31] | Point-based model | Postoperative | Pathology | 3 | Largest tumor diameter, bilobar extension, vascular invasion | HBsAg, HCV, tumor number, vascular invasion, tumor differentiation, cirrhosis |

AFP, alpha-fetoprotein; anti-Hbc, hepatitis B core antibodies; CIT, cold ischemia time; CK-19, cytokeratin-19; DWIT, donor warm ischemia time; ESLW, Estimated standard liver weight; GPC3, glypican-3; HBsAg, hepatitis B surface antigen; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; LT, liver transplantation; mRECIST, modified response evaluation criteria in solid tumors; MVI, microvascular invasion; NA, not available; NLR, neutrophil-lymphocyte ratio; TTV, total tumor volume; UCSF, University of California, San Francisco; WT, waiting time.

**Table 4.** Risk of bias and applicability concerns according to PROBAST.

| | ROB | | | | Applicability | | | Overall | |
|---|---|---|---|---|---|---|---|---|---|
| Study | Participants | Predictors | Outcome | Analysis | Participants | Predictors | Outcome | ROB | Applicability |
| Ma 2019 [14] | − | + | ? | − | − | + | ? | − | − |
| Shimamura 2019 [15] | + | + | ? | + | + | + | ? | ? | ? |
| Ling 2018 [16] | + | + | ? | − | + | + | ? | − | ? |
| Jiang 2018 [17] | − | ? | ? | − | − | ? | ? | − | − |
| Halazun 2017 [18] | + | + | + | + | + | + | + | + | + |
| Feng 2017 [19] | + | ? | ? | − | + | ? | ? | − | ? |
| Mehta 2016 [20] | + | + | + | + | + | + | + | + | + |
| Piñero 2016 [21] | − | + | ? | − | − | + | ? | − | ? |
| Lai 2016 [22] | + | + | + | + | + | + | + | + | + |
| Lee 2016 [23] | − | + | + | + | − | + | + | − | − |
| Agopian 2015 [24] | + | + | + | + | + | + | + | + | + |
| Li 2015 [25] | + | + | ? | − | + | + | ? | − | ? |
| Duvoux 2012 [26] | + | + | + | + | + | + | + | + | + |
| Wang 2011 [27] | + | + | ? | − | + | + | ? | − | ? |
| Decaens 2010 [28] | + | + | + | + | + | + | + | + | + |
| Chan 2008 [29] | − | + | + | + | − | + | + | − | − |
| Parfitt 2007 [30] | − | + | ? | − | − | + | ? | − | − |
| Iwatsuki 2000 [31] | − | + | ? | − | − | + | ? | − | − |
| Fernández 2018* [38] | − | + | ? | − | − | + | ? | − | − |
| Rhu 2018* [35] | + | + | ? | + | + | + | ? | ? | ? |
| Mehta 2018* [39] | + | + | + | + | + | + | + | + | + |
| Costentin 2017* [37] | + | + | ? | + | + | + | ? | ? | ? |
| Notarpaolo 2017* [34] | + | + | + | + | + | + | + | + | + |
| Piñero 2016* [33] | + | + | ? | + | + | + | ? | ? | ? |
| Varona 2015* [32] | − | + | ? | − | − | + | ? | − | − |
| Marelli 2008* [36] | − | + | ? | − | − | + | ? | − | − |

−, high ROB/high concern regarding applicability; ?, unclear ROB/unclear concern regarding applicability; +, low ROB/low concern regarding applicability; PROBAST, Prediction model Risk Of Bias ASsessment Tool; ROB, risk of bias.

*Only validation study.

discriminatory performance. For that, further validation of these risk models is mandatory, and internal validation should be conducted at least by using bootstrapping techniques. The quality assessment using the PROBAST tool also showed that only six original risk models had a good quality with low ROB and applicability concerns; however, these models were derived from populations where HCV is the predominant cause of HCC, and there was a lack of independent external validation, with the only exception being the AFP model (Table 5). It is noteworthy that the risk of HCC after LT is strongly associated with the HBV viral load. However, there is still controversy about the effect of HCV [47,48]. Although some risk models involve easy-to-measure factors, they are still complex and impractical. For example, the nomogram of Agopian et al. [24] consists of seven risk factors, making it hard to calculate and limiting its clinical use, regardless of the perfect discriminatory performance (AUC = 0.85). The risk prediction model should also be derived from a representative sample size to effectively reflect actual clinical practice. Our review showed that three prediction models were based on a small sample size (<100 participants) [21,29,30]. Even if these models showed good discrimination, their results should be interpreted with caution, because a small sample size can maximize sampling error and does not reflect clinical reality. So the design of any future risk prediction model must balance all these factors.

The AFP model was developed in a French cohort of 537 patients and was validated internally and externally in four studies [32–35], with a total of 1845 patients distributed among diverse populations and having different HCC causes. This made it the best-validated prediction model, with a good discriminatory performance and overall quality. NRI calculation was performed either in the original or in the validation studies to evaluate the ability of the AFP model to identify patients

**Table 5.** Recommendations of the well reported and well-validated risk prediction models.

| Model | Study type | Predominant Cause of HCC | Overall quality according to PROBAST | Model type | Performance* | TRIPOD level† | Independent external validation in HBV areas (yes/no) | Recommendation |
|---|---|---|---|---|---|---|---|---|
| Halazun 2017 [18] | Prospective | HCV | + | Preoperative<br>Postoperative<br>General | 0.82 (pre)<br>0.88 (post)<br>0.91 (general) | 1a | No | Not ready for clinical use and validation is required |
| Mehta 2016 [20] | Retrospective | '' | + | General | 0.77 (D)<br>0.82 (V)<br>0.75 (EV) | 3 | No | Recommended in HCV regions/independent external validation is required in HBV regions |
| Lai 2016 [22] | '' | '' | + | General | 0.58 (D)<br>0.59 (V) | 3 | No | Not ready for clinical use and validation is required |
| Agopian 2015 [24] | '' | '' | + | General | 0.79 (pre)<br>0.85 (post) | 3 | No | Not ready for clinical use and validation is required |
| Duvoux 2012 [26] | '' | '' | + | Preoperative | 0.70 | 3 | Yes | Recommended‡ |
| Decaens 2010 [28] | '' | '' | + | Preoperative | 0.65(D)<br>0.75(EV) | 3 | No | Recommended in HCV regions/independent external validation is required in HBV regions |
| Chan 2008 [29] | '' | '' | – | Postoperative | 0.91(D)<br>1.00 (V1)<br>0.95 (V2)<br>0.67(EV)<br>0.81(EV) | 3 | No | Not recommended‡ |
| Iwatsuki 2000 [31] | '' | '' | – | Postoperative | 0.71(EV) | 1a | No | Not recommended‡ |

–, high ROB/high concern regarding applicability; +, low ROB/low concern regarding applicability; '', same to the above cell; D, development cohort; EV, external validation; HBV, hepatitis B virus; HCV, hepatitis C virus; post, postoperative; pre, preoperative; PROBAST, Prediction model Risk Of Bias Assessment Tool.

*Refer to Tables 1 and 2 for other performance measures.

†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, nonrandom split-sample development and validation; 3, development and validation using separate data; 4, validation study.

‡Please refer to the Tables 1 and 2 and discussion for more details.

with high recurrence compared to Milan criteria, by reclassifying the recurrence and no recurrence subgroups between the Milan criteria and the AFP model. For instance, in the development cohort, the prediction of recurrence was improved significantly by using the AFP model compared with the Milan criteria (NRI = 0.11, $z = 3.28$, $P = 0.001$). However, in the validation cohort, there was no significant difference (NRI = 0.15, $z = 1.92$, $P = 0.055$), probably because of the smaller sample size with fewer events [26]. Likewise, in Notarpaolo et al.'s external validation study [34], by using NRI for recurrence at two years, there was a significant improvement of patients without recurrence compared to the Milan criteria (NRI = 0.14, $z = 6.81$, $P < 0.001$), but the total NRI was not significantly different (NRI = 0.03, $z = 0.43$, $P > 0.05$), because NRI for recurrence was similar for the Milan criteria and the AFP model. Given these good results, we feel that the AFP model deserves further prospective validation in diverse populations.

Generally, risk prediction models can be classified into three categories: preoperative, postoperative, and general models. Our review revealed that nine risk prediction models were based on preoperative data, five models were obtained from postoperative data, and six models were derived from both pre- and postoperative data, so they were considered general risk prediction models (Table 3).

The preoperative risk prediction models are mainly based on radiological and serological predictors, so they offer an adequate candidate selection and provide an approximate estimation of the future risk of developing HCC recurrence, thereby enhancing communication with patients and their relatives. To evaluate the risk of HCC recurrence after LT, we recommend that the assessment of preoperative risk models should be conducted in the context of other expanded selection criteria, which is out of the scope of this study. Also, relying only on pretransplant estimation is not a perfect approach, so it is mandatory to take into consideration both pre- and postoperative predictors, because an underestimation of more than 30% was observed between preoperative radiological and postoperative pathological assessments [20,33,37,49–52].

The postoperative risk prediction models are usually derived from pathological risk factors, while general risk prediction models incorporate both pre- and postoperative risk factors, so they cannot be used as a selection tool for HCC candidates for LT. However, these risk models can be utilized effectively after LT for different aspects, such as determining whether HCC surveillance after LT is necessary. For example, the RETREAT score is recommended to determine optimal screening intervals as well as to identify patients with a low risk of recurrence, so surveillance for this group is not required. This can reduce the cost and avoid the harmful effects of radiation and contrast [20,42].

The postoperative and general risk prediction models could also potentially affect the use of neoadjuvant therapies and immunosuppressants after LT, not only by identifying patients at high risk for HCC recurrence but also by providing a reference for the predicted probability of HCC recurrence, which is very helpful for designing future clinical trials. In other words, patients with a high risk of post-LT HCC recurrence should be considered for enrollment into clinical trials using post-LT neoadjuvant therapies to decrease the risk of recurrence, instead of waiting to treat patients after they have been already diagnosed with HCC recurrence.

Our study has several advantages. This study is the first systematic review that summarizes the discriminatory performance of the current risk models for the prediction of HCC recurrence after LT. Our search was performed with comprehensive approaches using precise inclusion criteria to identify the potential papers. Two independent authors collected data and evaluated the ROB in the involved studies. However, our review has some limitations, which should also be mentioned. First, the existing differences in study characteristics, populations, and risk factors in the prediction models may lead to high heterogeneity among the included studies. Second, a limited number of validation studies have been conducted to compare multiple risk prediction models in the same cohort. Third, the wide time range of the included studies (1981–2016), in which some novel diagnostic and therapeutic approaches for HCC (e.g., imaging, direct antiviral therapy) have been introduced, probably had a direct impact on survival and recurrence after LT. Fourth, although we have tried to extract all relevant data from the included studies, some missing information was unavoidable. Finally, most risk prediction models were developed based on specific population characteristics, which might not apply to other populations. Thus, these risk models should be validated externally in different populations.

In summary, we found 18 original risk prediction models intended to predict HCC recurrence in patients who had undergone LT. The quality of the studies was generally low, according to GRADE criteria. The most common included risk factors were AFP, tumor size, vascular invasion, tumor number, tumor differentiation, and NLR. Only five risk prediction models were

externally validated. At least eight prediction models showed good discriminatory performance in internal or external validation. The AFP model was the well-validated preoperative risk model that could stratify patients into high- and low-risk groups. Prospective and independent external validation of the current risk prediction models in diverse populations is highly recommended.

## Funding

This work supported by the National S&T Major Project (Grant number: 2017ZX10203205); Medical Science and Technology Project of Zhejiang Province (Grant number: 2016146968) and China Postdoctoral Science Foundation (Grant number: 2017M612014).

## Conflicts of interest

The authors have declared no conflicts of interest.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Table S1.** PRISMA 2009 checklist

**Data S1.** Search strategies

## REFERENCES

1. Villanueva A. Hepatocellular carcinoma. *N Engl J Med* 2019; **380**: 1450.
2. Sapisochin G, Bruix J. Liver transplantation for hepatocellular carcinoma: outcomes and novel surgical approaches. *Nat Rev Gastroenterol Hepatol* 2017; **14**: 203.
3. Sharma P, Welch K, Hussain H, *et al.* Incidence and risk factors of hepatocellular carcinoma recurrence after liver transplantation in the MELD era. *Dig Dis Sci* 2012; **57**: 806.
4. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339**: b2535.
5. 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 Medicine* 2014; **11**: e1001744.
6. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350**: g7594.
7. Atkins D, Best D, Briss PA, *et al.* Grading quality of evidence and strength of recommendations. *BMJ* 2004; **328**: 1490.
8. 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**: 51.
9. Pavlou M, Ambler G, Seaman SR, *et al.* How to develop a more accurate risk prediction model when there are few events. *BMJ* 2015; **351**: h3868.
10. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; **338**: b605.
11. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol* 2010; **172**: 971.
12. Debray TP, Damen JA, Snell KI, *et al.* A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017; **356**: i6460.
13. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med* 2014; **160**: 122.
14. Ma KW, She WH, Chan ACY, *et al.* Validated model for prediction of recurrent hepatocellular carcinoma after liver transplantation in Asian population. *World J Gastrointest Oncol* 2019; **11**: 322.
15. Shimamura T, Akamatsu N, Fujiyoshi M, *et al.* Expanded living-donor liver transplantation criteria for patients with hepatocellular carcinoma based on the Japanese nationwide survey: the 5-5-500 rule – a retrospective study. *Transplant Int* 2019; **32**: 356.
16. Ling Q, Liu J, Zhuo J, *et al.* Development of models to predict early post-transplant recurrence of hepatocellular carcinoma that also integrate the quality and characteristics of the liver graft: a national registry study in China. *Surgery* 2018; **164**: 155.
17. Jiang N, Zeng KN, Dou KF, *et al.* Preoperative alfa-fetoprotein and fibrinogen predict hepatocellular carcinoma recurrence after liver transplantation regardless of the Milan criteria: model development with external validation. *Cell Physiol Biochem* 2018; **48**: 317.
18. Halazun KJ, Najjar M, Abdelmessih RM, *et al.* Recurrence after liver transplantation for hepatocellular carcinoma: a new MORAL to the story. *Ann Surg* 2017; **265**: 557.
19. Feng J, Wu J, Zhu R, *et al.* Simple risk score for prediction of early recurrence of hepatocellular carcinoma within the Milan criteria after orthotopic liver transplantation. *Sci Rep* 2017; **7**: 44036.
20. Mehta N, Heimbach J, Harnois DM, *et al.* Validation of a risk estimation of tumor recurrence after transplant (RETREAT) score for hepatocellular carcinoma recurrence after liver transplant. *JAMA Oncol* 2017; **3**: 493.
21. Pinero F, Marciano S, Anders M, *et al.* Identifying patients at higher risk of hepatocellular carcinoma recurrence after liver transplantation in a multicenter cohort study from Argentina. *Eur J Gastro Hepatol* 2016; **28**: 421.
22. Lai Q, Nicolini D, Inostroza Nunez M, *et al.* A novel prognostic index in patients with hepatocellular cancer waiting for liver transplantation: time-radiological-response-alpha-fetoprotein-inflammation (TRAIN) score. *Ann Surg* 2016; **264**: 787.
23. Lee JH, Cho Y, Kim HY, *et al.* Serum tumor markers provide refined prognostication in selecting liver transplantation candidate for

hepatocellular carcinoma patients beyond the Milan criteria. *Ann Surg* 2016; **263**: 842.
24. Agopian VG, Harlander-Locke M, Zarrinpar A, *et al*. A novel prognostic nomogram accurately predicts hepatocellular carcinoma recurrence after liver transplantation: analysis of 865 consecutive liver transplant recipients. *J Am Coll Surg* 2015; **220**: 416.
25. Li Y, Ruan DY, Yi HM, Wang GY, Yang Y, Jiang N. A three-factor preoperative scoring model predicts risk of recurrence after liver resection or transplantation in hepatocellular carcinoma patients with preserved liver function. *Hepatobiliary Pancreat Dis Int* 2015; **14**: 477.
26. Duvoux C, Roudot-Thoraval F, Decaens T, *et al*. Liver transplantation for hepatocellular carcinoma: a model including alpha-fetoprotein improves the performance of Milan criteria. *Gastroenterology* 2012; **143**: 986; quiz e14-5.
27. Wang GY, Yang Y, Li H, *et al*. A scoring model based on neutrophil to lymphocyte ratio predicts recurrence of HBV-associated hepatocellular carcinoma after liver transplantation. *PLoS ONE* 2011; **6**: e25295.
28. Decaens T, Roudot-Thoraval F, Badran H, *et al*. Impact of tumour differentiation to select patients before liver transplantation for hepatocellular carcinoma. *Liver Int* 2011; **31**: 792.
29. Chan EY, Larson AM, Fix OK, *et al*. Identifying risk for recurrent hepatocellular carcinoma after liver transplantation: implications for surveillance studies and new adjuvant therapies. *Liver Transplant* 2008; **14**: 956.
30. Parfitt JR, Marotta P, Alghamdi M, *et al*. Recurrent hepatocellular carcinoma after transplantation: use of a pathological score on explanted livers to predict recurrence. *Liver Transplant* 2007; **13**: 543.
31. Iwatsuki S, Dvorchik I, Marsh JW, *et al*. Liver transplantation for hepatocellular carcinoma: a proposal of a prognostic scoring system. *J Am Coll Surg* 2000; **191**: 389.
32. Varona MA, Soriano A, Aguirre-Jaime A, *et al*. Risk factors of hepatocellular carcinoma recurrence after liver transplantation: accuracy of the alpha-fetoprotein model in a single-center experience. *Transpl Proc* 2015; **47**: 84.
33. Pinero F, Tisi Bana M, de Ataide EC, *et al*. Liver transplantation for hepatocellular carcinoma: evaluation of the alpha-fetoprotein model in a multicenter cohort from Latin America. *Liver Int* 2016; **36**: 1657.
34. Notarpaolo A, Layese R, Magistri P, *et al*. Validation of the AFP model as a predictor of HCC recurrence in patients with viral hepatitis-related cirrhosis who had received a liver transplant for HCC. *J Hepatol* 2017; **66**: 552.
35. Rhu J, Kim JM, Choi GS, Kwon CHD, Joh JW. Validation of the alpha-fetoprotein model for hepatocellular carcinoma recurrence after transplantation in an Asian population. *Transplantation* 2018; **102**: 1316.
36. Marelli L, Grasso A, Pleguezuelo M, *et al*. Tumour size and differentiation in predicting recurrence of hepatocellular carcinoma after liver transplantation: external validation of a new prognostic score. *Ann Surg Oncol* 2008; **15**: 3503.
37. Costentin CE, Amaddeo G, Decaens T, *et al*. Prediction of hepatocellular carcinoma recurrence after liver transplantation: comparison of four explant-based prognostic models. *Liver Int* 2017; **37**: 717.
38. Miron Fernandez I, Leon Diaz FJ, Sanchez Segura J, *et al*. Comparison of 3 explant-based prognostic models as predictors of hepatocellular carcinoma recurrence after liver transplantation: analysis of our experience. *Transpl Proc* 2019; **51**: 80.
39. Mehta N, Dodge JL, Roberts JP, Yao FY. Validation of the prognostic power of the RETREAT score for hepatocellular carcinoma recurrence using the UNOS database. *Am J Transplant* 2018; **18**: 1206.
40. Mazzaferro V, Regalia E, Doci R, *et al*. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med* 1996; **334**: 693.
41. Mazzaferro V, Bhoori S, Sposito C, *et al*. Milan criteria in liver transplantation for hepatocellular carcinoma: an evidence-based analysis of 15 years of experience. *Liver Transplant* 2011; **17**(Suppl 2): S44.
42. 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**: 723.
43. Yao FY, Ferrell L, Bass NM, *et al*. Liver transplantation for hepatocellular carcinoma: expansion of the tumor size limits does not adversely impact survival. *Hepatology* 2001; **33**: 1394.
44. Zheng SS, Xu X, Wu J, *et al*. Liver transplantation for hepatocellular carcinoma: Hangzhou experiences. *Transplantation* 2008; **85**: 1726.
45. Mazzaferro V, Llovet JM, Miceli R, *et al*. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. *Lancet Oncol* 2009; **10**: 35.
46. Xu X, Lu D, Ling Q, *et al*. Liver transplantation for hepatocellular carcinoma beyond the Milan criteria. *Gut* 2016; **65**: 1035.
47. Bozorgzadeh A, Orloff M, Abt P, *et al*. Survival outcomes in liver transplantation for hepatocellular carcinoma, comparing impact of hepatitis C versus other etiology of cirrhosis. *Liver Transplant* 2007; **13**: 807.
48. Wu TJ, Chan KM, Chou HS, *et al*. Liver transplantation in patients with hepatitis B virus-related hepatocellular carcinoma: the influence of viral characteristics on clinical outcome. *Ann Surg Oncol* 2013; **20**: 3582.
49. Sapisochin G, Goldaracena N, Laurence JM, *et al*. The extended Toronto criteria for liver transplantation in patients with hepatocellular carcinoma: a prospective validation study. *Hepatology* 2016; **64**: 2077.
50. Otto G, Schuchmann M, Hoppe-Lotichius M, *et al*. How to decide about liver transplantation in patients with hepatocellular carcinoma: size and number of lesions or response to TACE? *J Hepatol* 2013; **59**: 279.
51. Freeman RB, Mithoefer A, Ruthazer R, *et al*. Optimizing staging for hepatocellular carcinoma before liver transplantation: a retrospective analysis of the UNOS/OPTN database. *Liver Transplant* 2006; **12**: 1504.
52. Silva MF, Sherman M. Criteria for liver transplantation for HCC: what should the limits be? *J Hepatol* 2011; **55**: 1137.