AHBPS

Annals of Hepato-Biliary-Pancreatic Surgery

Check for updates

Review Article

Scoring systems for the management of oncological hepato-pancreato-biliary patients

Alexander W. Coombs*, Chloe Jordan*, Sabba A. Hussain, Omar Ghandour

Department of Surgery and Cancer, Imperial College London, London, United Kingdom

Oncological scoring systems in surgery are used as evidence-based decision aids to best support management through assessing prognosis, effectiveness and recurrence. Currently, the use of scoring systems in the hepato-pancreato-biliary (HPB) field is limited as concerns over precision and applicability prevent their widespread clinical implementation. The aim of this review was to discuss clinically useful oncological scoring systems for surgical management of HPB patients. A narrative review was conducted to appraise oncological HPB scoring systems. Original research articles of established and novel scoring systems were searched using Google Scholar, PubMed, Cochrane, and Ovid Medline. Selected models were determined by authors. This review discusses nine scoring systems in cancers of the liver (CLIP, BCLC, ALBI Grade, RETREAT, Fong's score), pancreas (Genç's score, mGPS), and biliary tract (TMHSS, MEGNA). Eight models used exclusively objective measurements to compute their scores while one used a mixture of both subjective and objective inputs. Seven models evaluated their scoring performance in external populations, with reported discriminatory c-statistic ranging from 0.58 to 0.82. Selection of model variables was most frequently determined using a combination of univariate and multivariate analysis. Calibration, another determinant of model accuracy, was poorly reported amongst nine scoring systems. A diverse range of HPB surgical scoring systems may facilitate evidence-based decisions on patient management and treatment. Future scoring systems need to be developed using heterogenous patient cohorts with improved stratification, with future trends integrating machine learning and genetics to improve outcome prediction.

Key Words: Models, statistical; Decision support techniques; Neoplasms

INTRODUCTION

Hepato-pancreato-biliary (HPB) surgery comprises treatment of benign and malignant diseases of the liver, pancreas and biliary tract [1]. Primary HPB cancers are relatively uncommon compared to other malignancies, although their incidence is increasing worldwide [2]. HPB surgeons commonly need to perform oncological treatment for colorectal cancer (CRC) liver metastases, with approximately half of all CRC patients developing metastatic liver disease [3]. Despite improvements in cancer surveillance and advances in treatment, HPB malignancies are still associated with high mortality rates [4]. This is partly due to their atypical presentation, meaning that patients are often diagnosed with an advanced disease. The principal management choice of these patients is surgical resection, although such surgeries are complex with high complication rates [4]. Additionally, as HPB malignancies are most prevalent in elderly populations, surgery potentially proposes a plethora of challenges. Therefore, adequate patient selection is essential [5]. Here, scoring systems for patients who are likely to benefit from surgical intervention and for patients who should be treated more conservatively may positively support a HPB surgeon's clinical decision-making.

Scoring systems in surgery have long been established to guide clinicians with a multitude already being utilised in regular clinical practice to predict surgical outcomes [6]. They can be applied for numerous reasons: to risk stratify patients into treatment groups, or to estimate an individual's risk of complications, prognosis, and recurrence [7,8]. Risk assessment

is imperative in clinical management. Evidence-based stratification tools compliment the shared decision-making process of both patients and clinicians [9]. This assessment can occur at any stage (pre-, peri-, or post-operative), although most scoring systems have been designed to be conducted prior to surgery [10].

An ideal scoring system should be accurate and simple using objective variables [11,12]. Notwithstanding the availability of various tools in the literature, they mostly fail to be implemented in regular clinical practice [13]. Low uptake can be ascribed to a lack of awareness of risk stratification models and concerns regarding their applicability, complexity and precision [11,12]. Additionally, the validation quality of scoring systems is non-uniform, with many models failing to meet the minimum requirements of study quality based on specialised quality metrics. This results in a substantial amount of research waste. In addition, often there is no clear evidence as to what clinical and economic effects would be of a given model [14,15]. Despite this, published scoring systems still offer clinical value with a culmination of effort being focused on increasing their predictive accuracies through the development of novel methods and refining existing models [10].

The aim of this review is to provide an overview of various HPB oncological scoring systems, assess the development quality of these systems, and evaluate performances of these models.

METHODOLOGY

We conducted a narrative review to provide an overview of potential clinical valuable and novel scoring systems that might be considered for use in the management of HPB oncological patients. Original research articles were identified using PubMed, MEDLINE (Ovid), Cochrane, and Google Scholar (Supplementary Material 1).

The scoring system chosen for review was based on current clinical endorsement and its validation in multiple populations, or alternatively if the score's function was deemed to be novel in HPB oncological management. When possible, scoring systems that had been externally validated were favourably selected. The effectiveness of each scoring system was primarily based on the c-statistic in the original and external study cohort populations. The c-statistic measures the ability of a scoring system to distinguish between those who do experience an outcome of interest and those who do not [16]. The measure of performance is also known as the discrimination ability. C-statistic scores lie between 0.5 and 1. A c-statistic of 0.5 infers no predictive ability (no better than random chance) and a c-statistic of 1 denotes perfect discrimination. In line with other previous reviews, a c-statistic > 0.7 was deemed 'acceptable' and a c-statistic > 0.8 was deemed 'good' discriminatory power [17,18]. Authors also assessed and reported calibration of scoring systems, another important measurement of model

performance. Calibration is an assessment of accurate alignment between estimated predicted risk and observed risk in models [19]. A scoring system might have good discrimination. However, if it is poorly calibrated, then the model's predictive ability will be inaccurate. This is especially important when assessing scoring systems that will ultimately be used in decision-making. Scoring systems were also subjectively analysed by authors for their applicability in clinical practice. Factors considered were the accessibility of variables used, number of variables needed and the use of objective measurements. A scoring system in this review was defined as a model or score that could risk-stratify patients into different categories to assist in the decision-making management of patients. At least two scoring systems for each organ in HPB were selected for variation purposes. Articles not in the English language were excluded. Abbreviations of selected scoring systems are summarised in Table 1.

SCORING SYSTEMS

Liver

Liver cancer is the second leading cause of all cancer related deaths globally, with hepatocellular carcinoma (HCC) being the most common variation (75%–85%) of primary liver cancer [2,20]. Robust and universal models that accurately assess prognosis, recurrence and function are valuable in assisting effective patient management.

Cancer of the Liver Italian Program (CLIP)

Mortality from HCC has progressively increased across Western Europe, making prognostic measures imperative in improving decision-making and outcomes [21]. CLIP is extensively used clinically as a prognostic tool for survival in HCC patients [22,23]. CLIP scores points based on the severity of certain factors in order to calculate survival as an endpoint (Table 2). Patients with higher scores indicate poorer prognosis. CLIP is the current standard for comparing effectiveness of new models. It is an improvement to Okuda, one of preliminary combined scoring systems in HCC [24].

CLIP has been externally validated in numerous populations

Table 1. Abbreviations of scoring systems

Abbreviation	Expanded term
ALBI	Albumin-Bilirubin
BCLC	Barcelona-Clinic Liver Cancer
CLIP	Cancer of the Liver Italian Program
MEGNA	Multifocality, Extra-hepatic Extension, Grade, Node Positivity, Age (> 60 yr) prognostic score
mGPS	Modified Glasgow Prognostic Score
RETREAT	Risk Estimation of Tumour Recurrence after Transplant
TMHSS	Tata Memorial Hospital Scoring System

Table 2. Variables used within each scoring system, their definition and outcome, and the statistical method of variable selection

Scoring system	Variable	Variable definition	Outcome measured by scoring system	Statistical method of variable selection
CLIP [21]	Child-Pugh score	A	Survival prognosis	Univariate and multivariate analysis
		B		
		C		
	Tumour morphology	Uninodular and extension ≤ 50%		
		Multinodular and extension ≤ 50%		
		Massive or extension ≥ 50%		
	Serum AFP	< 400 ng/dL		
		≥ 400 ng/dL		
	Portal vein thrombosis	Yes		
		No		
BCLC [38]	Tumour stage	1 HCC < 2 cm	Staging and treatment	Multivariate analysis and literature review
		1 HCC < 5 cm or 3 nodules < 3 cm		
		Multinodular		
		Portal invasion, N1 M1		
		Any		
	Child-Pugh stage	A		
		B		
		C		
	ECOG performance status	0		
		1–2		
		> 2		
ALBI [44]	Serum albumin	Nomogram	Liver function	Univariate and multivariate analysis
	Serum bilirubin	Nomogram		
RETREAT [51]	Serum AFP at liver transplantation (ng/mL)	0–20	Recurrence	Univariate and multivariate analysis
		21–99		
		100–999		
		≥ 1,000		
	Microvascular Invasion	Yes		
		No		
	Largest viable tumour diameter (cm) plus number of viable tumours	0		
		1.1–4.9		
		5.0–9.9		
		≥ 10		
Fong et al. [61]	Nodal invasion	Yes	Recurrence	Univariate and multivariate analysis
		No		
	Length of disease-free interval (mon)	≥ 12		
		< 12		
	Number of tumours	> 1		
		≤ 1		
	Tumour size (cm)	> 5		
		≤5		
	CEA (ng/mL)	> 200		
		≤ 200		
Genç et al. [69]	Tumour grade	1	Recurrence	Multivariate analysis
		2		
	Node positivity	Yes		
		No		
	Perineural invasion	Yes		
		No		

Table 2. Continued

Scoring system	Variable	Variable definition	Outcome measured by scoring system	Statistical method of variable selection
mGPS [72]	CRP	> 10	Survival	Univariate analysis
		< 10		
	Serum Albumin	> 35		
		< 35		
TMHSS [82]	Serum Bilirubin (mg/dL)	< 3	Management strategy	Univariate analysis
		> 3		
	Carbohydrate antigen 19-9 (U/mL)	0–30		
		30–90		
		90–450		
		> 450		
	Computed tomography scan	Normal		
		Gallbladder mass		
		Liver infiltration		
		Medially placed mass/intrahepatic biliary radicle dilatation		
		Metastatic disease		
MEGNA [87]	Multifocality	Yes	Survival	Multivariate analysis
		No		
	Extra-hepatic organ involvement	Yes		
		No		
	Tumour grade	Yes		
		No		
	Node positivity	Yes		
		No		
	Age > 60 yr	Yes		
		No		

CLIP, Cancer of the Liver Italian Program; BCLC, Barcelona Clinic Liver Cancer; ALBI, Albumin-Bilirubin Grade; RETREAT, Risk Estimation of Tumor Recurrence after Transplant Score; mGPS, Modified Glasgow Prognostic Score; TMHSS, Tata Memorial Hospital Scoring System; MEGNA, Multifocality, Extra-hepatic Extension, Grade, Node Positivity, Age (> 60 yr) prognostic score; AFP, alpha-fetoprotein; ECOG, Eastern Cooperative Oncology Group; CEA, carcinoembryonic antigen; CRP, C-reactive protein.

globally. It is cited as one of the most accurate tools in calculating HCC prognosis [25-27]. In patients with advanced HCC, CLIP is superior in predicting prognosis to other prominent systems. It has a better discriminative c-statistic (0.806) than the Japanese Integrated System (0.754) for example (Table 3) [26,28]. Other benefits include its simplicity due to the use of routinely available measures.

However, one variable, tumour morphology, has a subjective determinant of severity, thereby reducing the reliability of the score between users. Selection bias existed in the initial cohort as the patient group consisted mainly of middle-aged males [22]. Since then, it has been validated in numerous countries, thus increasing its global applicability [28-30]. The suitability of CLIP for patients undergoing surgical resection was also questioned, as most patients did not receive this treatment in the initial cohort [22]. There is also a limited range of risk stratification, as 80% of patients have a score of 0–2. Therefore, CLIP's use may only be beneficial for patients with higher scores [23].

Modifications to CLIP have recently been proposed, reporting improved prognostic prediction in the long term (c-statistic: modified CLIP-2 0.879; original CLIP 0.762) [31,32]. However, these modifications have yet to be clinically validated.

Barcelona Clinic Liver Cancer (BCLC)
The BCLC score is another prognostic tool used preoperatively to determine HCC staging and to risk-stratify patients into treatment groups [33]. These stages range from very early (0) to early (A), intermediate (B), advanced (C), and end-stage disease (D) (Table 2).

BCLC can provide both staging classification and treatment recommendations, thus increasing its clinical utility. These recommendations include surgical resection, transplantation or ablation (0 & A); transarterial chemoembolisation (TACE) (B); medical treatments (C); and palliative care (D). BCLC is also the most extensively used HCC score clinically in the western world due to its practicality and validation in western popula-

Table 3. Overview of the nine scoring systems explored in this review, including key components essential for their development and statistical analysis

Name	Year	Organ	Country	Type of study	Development cohort size	Development cohort sample method	Testing cohort (external validation) size	Testing cohort (external validation) sample method	Internal cohort (C-index)	External cohort (C-index)	Calibration metric score	Statistical methods
CLIP [21]	1998	Liver	Italy	RC	435	Split-sample validation	196	Same centre, different population	N/A	0.806	N/A	MCRA
BCLC [38]	1999	Liver	Spain	LR	N/A	N/A	766	Different centre	N/A	0.72	N/A	MRCA
ALBI [44]	2015	Liver	Japan	RC	1,313	Split-sample validation	5,097	Same and two additional centre, different population	N/A	0.61–0.68	N/A	MCRA
RETREAT [51]	2017	Liver	USA	RC	721	N/A	340	Different centre	0.77	0.82	N/A	MCRA
Fong et al. [61]	1999	Liver	USA	RC	1,001	N/A	N/A	N/A	N/A	0.62	N/A	MCRA
Genç et al. [69]	2018	Pancreas	Netherlands	RC	211	N/A	N/A	N/A	0.81	N/A	Hosmer–Lemeshow chi-square 11.25, $p = 0.258$	MCRA
mGPS [72]	2007	Pancreas	UK	RC	316	N/A	807	Different centre	N/A	N/A	N/A	MCRA
TMHSS [82]	2017	Gall bladder	India	RC	124	Split-sample validation	N/A	N/A	0.75	N/A	N/A	MCRA
MEGNA [87]	2008	Biliary tract	USA	RC	275	Split-sample validation	417	Different population	N/A	0.58	N/A	MCRA

CLIP, Cancer of the Liver Italian Program; BCLC, Barcelona Clinic Liver Cancer; ALBI, Albumin–Bilirubin Grade; RETREAT, Risk Estimation of Tumor Recurrence after Transplant Score; mGPS, Modified Glasgow Prognostic Score; TMHSS, Tata Memorial Hospital Scoring System; MEGNA, Multifocality, Extra-hepatic Extension, Grade, Node Positivity, Age (> 60 yr) prognostic score; RC, retrospective cohort; LR, literature review; MCRA, multi-variable cox regression analysis; N/A, not available.

tions [24,34]. Many organisations have endorsed this system in their guidelines, including the American Associations for the Study of the Liver (AASLD) and European Associations for the Study of the Liver (EASL) [21,35]. Studies validating BCLC also suggest that it is the most suitable for accurately predicting prognosis in potential surgical patients, a benefit over CLIP [36].

However, the clinical validity of BCLC over the recommended treatment for stage B patients (TACE) is debated. The initial cohort for stage B was heterogenous in nature with varied HCC features. Hence, TACE may not always be clinically appropriate to perform in stage B patients [37,38]. Furthermore, the score is missing an important variable (cause of underlying disease) and comorbid conditions, all of which may influence staging and treatment, respectively [38]. To improve this, new refinements of BCLC with better prognostic prediction and treatment choices have been reported [24]. One example is an expanded BCLC tool through the inclusion of the Milan criteria (which assesses suitability for liver transplantation) [39]. However, more extensive clinical and external validation is required before such expanded criteria can have widespread clinical implementation.

Albumin-Bilirubin Grade (ALBI)

Chronic liver disease (CLD) is a significant risk factor for the development of HCC [40]. It has been estimated that 70%–90% of all HCC patients have some degrees of concurrent CLD [41]. As CLD is a competing cause of death in HCC patients, understanding its severity is useful for predicting liver function and survival [42]. Currently, clinicians use the Child-Pugh score to assess CLD in HCC patients. However, this score fails to distinguish between different severities of CLD, categorising the majority of patients into class B. Additionally, it includes two subjective measurements: ascites and hepatic encephalopathy [43]. Therefore, a new scoring system (ALBI grade) has been developed specifically to grade liver function in HCC patients [44]. ALBI utilises a nomogram to categorise patients into 3 grades. It uses two simple and objective measurements: serum albumin and bilirubin (Table 2).

The ALBI grade has been validated to predict overall survival in patients of each BCLC stage undergoing various treatment options. It has demonstrated the suitability of use in heterogenous cohorts globally. Its applicability to all stages of HCC, along with its simplicity, makes ALBI a useful tool clinically [45]. Additionally, to support its use in practice, ALBI has exhibited a strong correlation with indocyanine green clearance, a bedside test that estimates liver function [46]. The application of this scoring system is expected to be useful for risk stratifying HCC patients in the pre-operative stage. Grade 1 patients may be the most suitable to undergo intent-curative hepatic resection, whereas grade 2 or 3 patients are more suitable for liver transplantation or other less invasive treatment options such as ablation therapies [46].

More recently, it has been suggested that a post-operative ALBI score calculation can predict overall survival more accurately than the pre-operative version [47]. Therefore, its application post-operatively may supplement an updated prognostic prediction [47]. One pitfall of ALBI is that patients who are assigned a grade 2 demonstrate a wide range of hepatic functions, which may allocate some patients into treatment options that might not be optimal. This potential issue has been raised and addressed in a study that proposes a newly modified version of the original model called mALBI, which further sub-categorises patients into grade 2a and 2b [48]. Moreover, the ALBI grade has not yet been endorsed for clinical use. It is still under critical review despite the literature reporting its effectiveness as a predictor of liver function and outcome in HCC patients.

Risk Estimation of Tumor Recurrence after Transplant Score (RETREAT)

Risk of recurrence remains an issue for HCC patients undergoing liver transplantation. Those experiencing recurrence (20%) have a poor median survival of 12 months [49]. Both the USA and UK currently incorporate the Milan criteria, a criteria that considers tumour size, number of tumours and tumour invasion, into their pre-operative guidance for liver transplantation selection in HCC patients [49,50]. The issue with the Milan criteria is that well-established risk factors of recurrence are not considered [51,52]. RETREAT, a novel scoring system, has been developed to guide clinicians as to which patients are at risk of recurrence and consequently may need to be followed-up more regularly [51]. This externally validated scoring system uses three easily objective and accessible variables to calculate a score that predicts risk of recurrence within 1 and 5 years of treatment (Table 2). Authors of RETREAT suggest that those with a score ≥ 4 would benefit from an adjunct therapy [51].

Research supports that increased frequency of follow up can improve post-recurrence survival of patients [53]. This proposes that intervention of increased monitoring may decrease mortality as recurrent HCC can be discovered and treated earlier. A large study including all patients who underwent liver transplant in the US population between 2012 and 2014 demonstrated significant prognostic ability by using RETREAT, thus validating its use [54]. Since then, The American Association for the study of Liver Diseases has endorsed RETREAT in its report as a useful way to determine follow-up intervals [55].

The limitation of using RETREAT as a prognostic tool is that the measurement of AFP is required at the time of liver transplantation. Therefore, this scoring system cannot be used pre-operatively to determine whether to proceed with transplantation. However, many criteria and scoring tools are available for this purpose [50]. Other scoring systems similar to the post-operative nature of RETREAT include Agopian et al.'s [56] nomogram and post-MORAL [57]. Both scores have

greater prognostic power than RETREAT, with c-statistic of 0.88 and 0.85, respectively. Despite their impressive c-statistics, these findings should be interpreted with caution as these scoring systems have yet to be validated with external populations. Therefore, their use over RETREAT is not currently recommended. Moreover, in Asian populations, the SNAPP scoring system may offer improved risk of recurrence prediction than RETREAT which has been validated in western populations [58].

Fong's score: clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer

Liver resection remains the key curative treatment option for colorectal cancer liver metastases (CRLM) with a cure rate of 20.6% [59]. Of patients who undergo hepatic resection for CRLM, 60%–85% experience cancer recurrence [60]. Therefore, similar to liver transplantation for HCC, it is vital that suitable patients with CRLM are identified for intent-curative surgery and that those with higher risk are followed up more vigilantly. Fong's scoring system risk stratifies CRLM patients into low (1–2) and high (3–5) risks of recurrence (Table 2) [61]. Patients with a low risk (1–2) are advised to consider intent-curative resection. Patients with scores 3–4 should consider adjunct therapy with resection. Patients with a score of 5 are indicated to have very poor prognosis. Thus, extra consideration is required about the decision to surgically resect these patients.

Many prognostic scoring tools have been produced for curative liver resection of colorectal liver metastases in HPB oncology [62]. Currently, Fong's score is the most well-known and understood model in the literature. However, others have tried to eclipse its prognostic ability. In a meta-analysis, Fong was compared other scores externally validated in at least two study populations [62]. Fong's score was significant for predicting rate free survival (RFS), but not significant for estimating overall survival. For predicting RFS, the Basingstoke score had a greater prognostic power (c-statistic: 0.74) than Fong, although its confidence intervals (CI) were widely distributed (CI: 0.52–0.88). For estimating overall survival, clinicians may opt to use the Valentini nomogram which has a significant prognostic power (c-statistic: 0.71).

In consideration that a prognostic model c-statistic > 0.7 is 'acceptable', Fong's score does not meet this requirement. Therefore, it may not be as useful as first thought for identifying high-risk patients despite its popularity in the literature. In light of this, its clinical endorsement may be overlooked and future models instead may be requested. Future models may need to incorporate genetic markers and utilise machine learning (ML) or neural networks to improve its prognostic power in predicting recurrence in CRLM patients [62,63]. Notwithstanding this disappointing c-statistic, Fong's score may still be useful for estimating RFS in CRLM patients, especially as this score has been externally validated in four different populations, more than any of its comparators.

Pancreas

Pancreatic cancer (PC) is the seventh leading cause of cancer related mortality worldwide, totalling 4.5% of total cancer deaths in 2018 [2]. Despite complete surgical resection being the only curative treatment for PC, merely 10%–15% of patients are eligible due to most presenting with advanced disease [64]. For this reason, efforts focusing on developing tools that can identify suitable patients for surgery may be helpful in PC management.

Genç's score: scoring system to predict recurrent disease in grades 1 and 2 nonfunctional pancreatic neuroendocrine tumors

Although neuroendocrine tumours (NETS) are perceived to be rare, they are increasing in incidence annually, displaying a 6.4-fold increase over a 39-year period [65]. Non-functional pancreatic neuroendocrine tumours (P-NETs) account for 60%–90% of all P-NETs. They are often discovered incidentally on imaging [66,67]. For malignant and localised disease, surgical resection of the P-NET is required to stop distant metastases from developing [66,68]. Overall, surgical treatment provides a good prognosis for patients. However, for those who experience recurrence, the prognosis is poor. Therefore, a scoring system that can identify patients at risk of recurrence is needed for clinicians.

A scoring system by Genç et al. [69] was the first to be proposed for risk-stratifying patients with non-functional P-NETs. Genç's score uses three easily accessible variables to calculate a total score out of 88 points. The score can be used to categorise patients into low risk (< 24) and high risk (> 24) of recurrence. For high-risk patients, authors suggest the use of an adjunct be considered to improve their prognosis. In terms of overall survival, low risk and high-risk patients had a mean survival of 110.3 and 99.4 months respectively [69]. Whilst the mean survival of the two groups is statistically significant, it may be argued that the difference of 10 months is not substantial, with high-risk patients just above the cut off possibly having similar outcomes to low-risk patients. This is likely considering that a low-risk patient with a score of 24 has 16% chance of recurrence, while a high-risk patient with a score of 40 has 25% chance of recurrence. Therefore, the decision to use an adjunct may be more difficult for patients with scores of 40 or 48 than for those with scores of 64 or 88. This scoring system may be most useful in the decision of follow up intervals after surgical treatment to monitor high risk patients. Genç's score has only been validated in the development cohort despite it having a well-performing c-statistic of 0.81. The lack of simple guidance to predicting recurrence in patients with non-functional P-NETS makes this a novel and clinically useful tool. However, its endorsement for deciding treatment options is not advocated until the scoring system has been externally validated with other populations. For predicting tumour recurrence in functional P-NETS or in Asian populations, a similar scoring

system described by Zou et al. [70] (c-statistic: 0.81) may be a suitable alternative.

Modified Glasgow Prognostic Score (mGPS)

In addition to tumour biological factors that make up the tumour-node-metastasis (TNM) staging system, there is mounting evidence to support systemic inflammation as a critical and independent marker of prognosis in malignancy [71]. As a result, inflammation based prognostic scores have been developed and proposed for prognostication and predicting survival outcomes in cancer patients [71]. The mGPS is a prognostic score used to grade inflammatory responses of PC by combining both C-reactive protein (CRP) and serum albumin levels. The mGPS has scores between 0–2, with a higher score indicating a worse prognosis (Table 2).

The main advantage of the mGPS is that it utilises two commonly recorded objective measures, making it easy and simple to calculate in a clinical setting. In a meta-analysis of 20 studies, mGPS demonstrated significant overall prognostic power with hazard ratio (HR) of 1.50 for those categorised as high risk by this score versus those with a low risk [72]. However, most of these studies were conducted in eastern populations, with western population studies showing inconclusive findings (HR: 1.34, $p = 0.268$). Another significant pitfall of the mGPS is the possibility of receiving inaccurate results due to elevated CRP or albumin levels unrelated to PC and instead due to concurrent co-morbidities or infection. Such inaccuracies may ultimately influence clinical decision-making. Therefore, good clinical judgement is needed when using the model in patients with comorbid disease.

More recently, the CRP/albumin ratio has shown potential [73]. Using the same two simple markers, one study showed that the CRP/albumin ratio had a better discriminatory ability (0.70) than the mGPS (0.63), though it should be validated with data from multiple centres before any recommendation could be made for its use over mGPS [73]. Additionally, a scoring system with haemostatic markers has been developed for use in patients with advanced pancreatic patients to provide an estimated median survival time [74]. Its novel use of markers provides an interesting area for future research. However, like the CRP/albumin ratio, its prognostic ability should be tested with external populations. Overall, the mGPS provides a valuable pre-operative prognostic ability, although future novel scores with improved discrimination may eclipse its use in western populations.

Biliary

Biliary tract cancers primarily include three different entities: gallbladder cancer (GBC), intrahepatic cholangiocarcinoma (ICC) and extrahepatic cholangiocarcinoma (ECC) [75]. Although uncommon, these malignancies are highly fatal with a poor prognosis. Thus, scoring systems may provide importance in risk stratification of patients.

Tata Memorial Hospital Scoring System (TMHSS)

Although collectively rare, GBC is the most common malignancy of the biliary tract. It is often diagnosed at an advanced stage. It remains difficult to treat mainly due to the associated poor prognosis, with surgery being the only curative option [75-77]. Currently, all available scoring systems for GBC are based on histopathological specimen results following cholecystectomy [78,79]. Thus, clinicians are unable to predict surgical outcomes preoperatively to support their clinical management. A long-held concern of GBC treatment is the offering of resection to jaundiced patients [80]. This is because jaundice is strongly associated with advanced disease, high mortality and inoperability [80]. However, recent studies have successfully demonstrated that a small sub-population of these patients can undergo surgery with curative intent and achieve long-term survival post-operatively [81]. TMHSS, a new pre-operative scoring system, might be helpful to clinicians in identifying GBC patients most suitable for surgical intervention and equally aiding to avoid surgery in those where such invasive treatment offers no overall benefit. The score has been proposed based on radiological, biochemical and clinical features [82]. The aim of TMHSS is to offer an estimation of prognosis and to predict resectability of GBC: Group A (score 0–3) being managed surgically, Group B (score 4–6) managed with neoadjuvant options or staging laparoscopy, and Group C (score ≥ 7) managed palliatively (Table 2).

Statistical methods used to derive the scoring system were not acknowledged [82]. Additionally, in the development cohort, surgery was considered amenable for 58% of patients in comparison with the validation cohort where surgery was only considered for 32.5% of the population. Discrepancies between development and validation cohorts highlight the heterogeneity and that testing in a larger population group inclusive of various ethnicities would be appropriate.

Strengths of TMHSS include its ease of computation since its variables (CA 19-9 and serum bilirubin) are routinely assessed in practice. The score has good discriminative ability in identifying patients who would benefit from radical resection. However, the scoring system only appears to be beneficial for patients at the two extremes of the disease—Group A and Group C. The benefit of the score is less clear for Group B, for whom the scoring system fails to differentiate patients who require surgery or palliative care, an area that requires further exploration [83]. Nevertheless, TMHSS has the potential to decrease unnecessary surgical explorations and to guide patients to an appropriate management strategy, thereby offering some prognostication. The TMHSS is a precise, simple and easily performed test with a high negative predictive value, showing its value in predicting the resectability of GBC.

Multifocality, Extra-hepatic Extension, Grade, Node Positivity, Age (> 60) prognostic score (MEGNA)

ICC, a malignancy arising from biliary epithelium, is the sec-

ond most common primary liver tumour. It accounts for 10%–15% of hepatic malignancies [84,85]. The decision to manage patients surgically often relies on the perceived oncological benefit. Although multiple prognostic nomograms have been developed to predict survival following surgical resection, they have failed to be clinically integrated due to their complexity [86]. The required variables for these nomograms can only be obtained post-operatively. In addition, most nomograms are not validated in a population representative sample. The MEG-NA prognostic score was developed to provide a more accurate pre-operative estimation of survival prediction following resection of ICC [87]. The allocated score (0–5) risk stratifies patients into four groups, with a higher score indicating a worse prognosis.

MEGNA is based on five pre-operative variables, highlighting its simplicity (Table 2). Despite being developed on data from patients who undergo resection, MEGNA has the ability to inform decisions regarding management of ICC patients, although additional studies are required to evaluate its use pre-operatively. Strengths of this scoring system include external validation and superiority in predicting prognosis and survival to existing models. A comparison of the prognostic separation index between MEGNA and staging systems of the American Joint Committee on Cancer (AJCC) has demonstrated the superiority of MEGNA in patient survival following hepatectomy [88].

However, its drawback is that currently MENGA is only potentially generalisable to US population. Current literature highlights substantial heterogeneity in histopathological and genomic characteristics of ICC based on geography [87]. This heterogeneity is demonstrated in a study by Hahn et al. [88] who evaluated the use of the MEGNA score with a direct comparison of its prognostic value with the AJCC staging system, the most widely used one currently. This study found that the C-index was 0.58 for MEGNA and 0.61 for AJCC, thus demon-
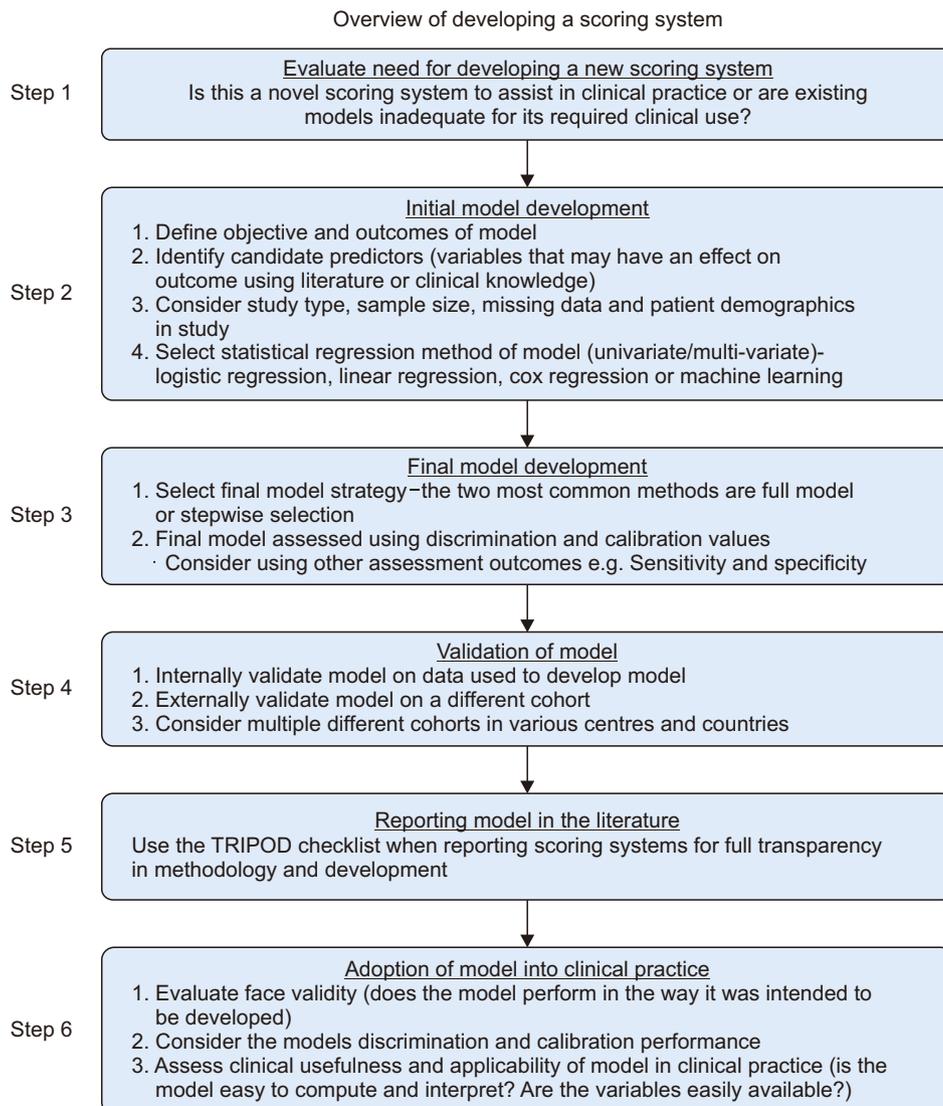
Overview of developing a scoring system

Step 1
Evaluate need for developing a new scoring system
Is this a novel scoring system to assist in clinical practice or are existing models inadequate for its required clinical use?

Step 2
Initial model development
1. Define objective and outcomes of model
2. Identify candidate predictors (variables that may have an effect on outcome using literature or clinical knowledge)
3. Consider study type, sample size, missing data and patient demographics in study
4. Select statistical regression method of model (univariate/multi-variate)-logistic regression, linear regression, cox regression or machine learning

Step 3
Final model development
1. Select final model strategy−the two most common methods are full model or stepwise selection
2. Final model assessed using discrimination and calibration values
  · Consider using other assessment outcomes e.g. Sensitivity and specificity

Step 4
Validation of model
1. Internally validate model on data used to develop model
2. Externally validate model on a different cohort
3. Consider multiple different cohorts in various centres and countries

Step 5
Reporting model in the literature
Use the TRIPOD checklist when reporting scoring systems for full transparency in methodology and development

Step 6
Adoption of model into clinical practice
1. Evaluate face validity (does the model perform in the way it was intended to be developed)
2. Consider the models discrimination and calibration performance
3. Assess clinical usefulness and applicability of model in clinical practice (is the model easy to compute and interpret? Are the variables easily available?)

Fig. 1. Flowchart illustrating the stepwise approach for developing a scoring system. TRIPOD, transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.

strating that the ability of the MEGNA score to predict individual patient prognosis was below the 0.7 threshold of acceptable. On the one hand, this study demonstrated that MEGNA was not superior to the AJCC system in this population. It also showed that neither AJCC nor MEGNA performed adequately to support clinical decisions within this population. Contrary to this, a multi-centre validation study by Schnitzbauer et al. [89] concluded that risk groups calculated by the MEGNA score resulted in significant stratification in survival, making it a good discriminator for ICC. In summary, current scoring systems for ICC are sub-optimal. Further research studies are needed to validate the utility of the MEGNA score across geographical populations [87].

Future trends

Efforts to find ideal oncological surgical scoring systems will continue, with developers striving for an objective, accurate, and economical tool. Future model development should thus seek to produce a tool using robust statistical models, increasing precision estimates whilst being externally validated with various heterogenous populations [90]. A basic overview of how model may be developed is presented in Fig. 1 [91-94]. Moreover, all future model development should be guided by and reported in accordance to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) guidelines to ensure methodological rigor and transparency [95].

Advancements in computing technologies and large data-set availability have led to the development of ML as an alternative approach to predict and stratify the risk of patient outcomes [96]. It has been posited to be advantageous over traditional regression-based scoring systems when there are large numbers of co-variates and non-linear relationships or when complex interactions are present. Although to date, ML studies have failed to show a performance benefit over traditional logistic regression models [97], data-driven technologies still hold potential advantages of being able to offer personalised care to patients from diagnosis, treatment, prognosis and monitoring in order to improve individual health outcomes. Concerns have been expressed regarding bias in ML, such as subjective selection of relevant variables using clinical knowledge that might influence results. Moreover, ML can only analyse data provided. Over-fitting of data may cause models to be unsuitable for other populations. In addressing the lack of validation and transparency of ML algorithms, a specific version of TRIPOD, the TRIPOD-ML, is being developed [98]. This checklist can provide methodological guidance on key items for researchers to consider and report during study development of future ML-based scoring systems, encouraging rigorous and transparent research in the ML community.

With the movement towards personalised medicine, newer scoring systems are being developed to incorporate genetics, such as the use of prognosis-associated genes in PC [97]. Mi-

croRNAs involved in gene regulation shows promising prognostic value due to their exclusivity to specific cell or tissue types [99]. Such genetic markers included in scoring systems could better predict outcomes and target individualised therapy for patients in the future.

Limitations

The main limitation of this review was the subjective nature by which the scoring systems were selected, with certain scoring systems of possible clinical use not being included. However, this subjectivity allowed authors to consider literature across the HPB speciality. The selected scoring systems were economical to use and reputable. Some provided novel use in the speciality. A further limitation was the use of the c-statistic to assess suitability for clinical application. Such models may be best assessed using other additional statistics and tools such as sensitivity and specificity [16,100]. Additionally, the used reference ranges were not completely justified based on an acceptable c-statistic as c-statistics close to the 0.7 mark may need caution [16]. The reporting of calibration was also poorly observed in this review. Therefore, while discrimination was satisfactory for most, a poor calibration result could weaken the use of such scores in clinical practice.

CONCLUSION

HPB oncological scoring systems discussed in this review provide a diverse overview of available models to predict prognosis, recurrence and function. These scoring systems may be considered for use in clinical environments to assist in the management of HPB patients. This could facilitate evidence-based discussions with patients so that they can make an informed and personally correct decision about their treatment. Current weaknesses of scoring systems include the lack of validation across heterogenous population groups, the non-reporting of calibration and the non-inclusion of various statistics to determine their clinical suitability. Scoring systems of the future should be designed and reported in consideration of TRIPOD. Personalised medicine and artificial intelligence will most likely become the trend of future surgical scoring systems.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found at https://doi.org/10.14701/ahbps.21-113.

ACKNOWLEDGEMENTS

FUNDING

CONFLICT OF INTEREST

ORCID

Alexander W. Coombs, https://orcid.org/0000-0002-5939-8960
Chloe Jordan, https://orcid.org/0000-0003-4827-5882
Sabba A. Hussain, https://orcid.org/0000-0001-7165-8609
Omar Ghandour, https://orcid.org/0000-0002-0009-7565

AUTHOR CONTRIBUTIONS

Conceptualization: All authors. Data curation: All authors. Methodology: All authors. Writing - original draft: All authors. Writing - review & editing: AWC, CJ.

REFERENCES

1. Oxford University Hospitals. HPB surgery [Internet]. Oxford: Oxford University Hospitals 2021 [cited 2020 Jan 11]. Available from: https://www.ouh.nhs.uk/services/departments/general-surgery/hpb-surgery/default.aspx.

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.

3. Cirocchi R, Trastulli S, Boselli C, Montedori A, Cavaliere D, Parisi A, et al. Radiofrequency ablation in the treatment of liver metastases from colorectal cancer. Cochrane Database Syst Rev 2012;(6):CD006317.

4. Alberts SR, Gores GJ, Kim GP, Roberts LR, Kendrick ML, Rosen CB, et al. Treatment options for hepatobiliary and pancreatic cancer. Mayo Clin Proc 2007;82:628-637.

5. Papis D, Vagliasindi A, Maida P. Hepatobiliary and pancreatic surgery in the elderly: current status. Ann Hepatobiliary Pancreat Surg 2020;24:1-5.

6. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ 2009;338:b604.

7. Jones HJ, de Cossart L. Risk scoring in surgical patients. Br J Surg 1999;86:149-157.

8. Barnett S, Moonesinghe SR. Clinical risk scores to guide perioperative management. Postgrad Med J 2011;87:535-541.

9. Knops AM, Legemate DA, Goossens A, Bossuyt PM, Ubbink DT. Decision aids for patients facing a surgical treatment decision: a systematic review and meta-analysis. Ann Surg 2013;257:860-866.

10. Chandra A, Mangam S, Marzouk D. A review of risk scoring systems utilised in patients undergoing gastroin-test Surg 2009;13:1529-1538.

11. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016;353:i3140.

12. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606.

13. Liao L, Mark DB. Clinical prediction models: are we building better mousetraps? J Am Coll Cardiol 2003;42:851-853.

14. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. BMJ 2019;367:l5358.

15. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328.

16. Caetano SJ, Sonpavde G, Pond GR. C-statistic: a brief explanation of its construction, interpretation and limitations. Eur J Cancer 2018;90:130-132.

17. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, et al. Risk prediction models for hospital readmission: a systematic review. JAMA 2011;306:1688-1698.

18. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, Lacaille D. A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol 2015;68:3-14.

19. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230.

20. Vogel A, Cervantes A, Chau I, Daniele B, Llovet JM, Meyer T, et al. Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2018;29(Suppl 4):iv238-iv255.

21. European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2018;69:182-236.

22. The Cancer of the Liver Italian Program (Clip) Investigators. A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients: the Cancer of the Liver Italian Program (CLIP) investigators. Hepatology 1998;28:751-755.

23. Kudo M, Chung H, Osaki Y. Prognostic staging system for hepatocellular carcinoma (CLIP score): its value and limitations, and a proposal for a new staging system, the Japan Integrated Staging Score (JIS score). J Gastroenterol 2003;38:207-215.

24. Kinoshita A, Onoda H, Fushiya N, Koike K, Nishino H, Tajiri H. Staging systems for hepatocellular carcinoma: current status and future perspectives. World J Hepatol 2015;7:406-424.

25. Llovet JM, Bruix J. Prospective validation of the Cancer of the Liver Italian Program (CLIP) score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. Hepatology 2000;32:679-680.

26. Chen ZH, Hong YF, Lin J, Li X, Wu DH, Wen JY, et al. Validation and ranking of seven staging systems of hepatocellular carcinoma. Oncol

Lett 2017;14:705-714.

27. Liu PH, Hsu CY, Hsia CY, Lee YH, Su CW, Huang YH, et al. Prognosis of hepatocellular carcinoma: assessment of eleven staging systems. J Hepatol 2016;64:601-608.

28. Huitzil-Melendez FD, Capanu M, O'Reilly EM, Duffy A, Gansukh B, Saltz LL, et al. Advanced hepatocellular carcinoma: which staging systems best predict prognosis? J Clin Oncol 2010;28:2889-2895.

29. Ueno S, Tanabe G, Sako K, Hiwaki T, Hokotate H, Fukukura Y, et al. Discrimination value of the new western prognostic system (CLIP score) for hepatocellular carcinoma in 662 Japanese patients. Cancer of the Liver Italian Program. Hepatology 2001;34:529-534.

30. Levy I, Sherman M. Staging of hepatocellular carcinoma: assessment of the CLIP, Okuda, and Child-Pugh staging systems in a cohort of 257 patients in Toronto. Gut 2002;50:881-885.

31. Huo TI, Huang YH, Lin HC, Wu JC, Chiang JH, Lee PC, et al. Proposal of a modified Cancer of the Liver Italian Program staging system based on the model for end-stage liver disease for patients with hepatocellular carcinoma undergoing loco-regional therapy. Am J Gastroenterol 2006;101:975-982.

32. Nanashima A, Morino S, Yamaguchi H, Tanaka K, Shibasaki S, Tsuji T, et al. Modified CLIP using PIVKA-II for evaluating prognosis after hepatectomy for hepatocellular carcinoma. Eur J Surg Oncol 2003;29:735-742.

33. Llovet JM, Brú C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999;19:329-338.

34. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet 2018;391:1301-1314.

35. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020-1022.

36. Vitale A, Saracino E, Boccagni P, Brolese A, D'Amico F, Gringeri E, et al. Validation of the BCLC prognostic system in surgical hepatocellular cancer patients. Transplant Proc 2009;41:1260-1263.

37. Barman PM, Sharma P, Krishnamurthy V, Willatt J, McCurdy H, Moseley RH, et al. Predictors of mortality in patients with hepatocellular carcinoma undergoing transarterial chemoembolization. Dig Dis Sci 2014;59:2821-2825.

38. Barman PM, Su GL. Limitations of the barcelona clinic liver cancer staging system with a focus on transarterial chemoembolization as a key modality for treatment of hepatocellular carcinoma. Clin Liver Dis (Hoboken) 2016;7:32-35.

39. Wang YY, Zhong JH, Xu HF, Xu G, Wang LJ, Xu D, et al. A modified staging of early and intermediate hepatocellular carcinoma based on single tumour >7 cm and multiple tumours beyond up-to-seven criteria. Aliment Pharmacol Ther 2019;49:202-210.

40. Tsukuma H, Hiyama T, Tanaka S, Nakao M, Yabuuchi T, Kitamura T, et al. Risk factors for hepatocellular carcinoma among patients with chronic liver disease. N Engl J Med 1993;328:1797-1801.

41. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 2007;132:2557-2576.

42. Okuda H. Hepatocellular carcinoma development in cirrhosis. Best Pract Res Clin Gastroenterol 2007;21:161-173.

43. Johnson P, Berhane S, Satomura S, Tada T, Kumada T, Teng M, et al. O110 An international collaborative study assessing the role of

aetiology and stage in survival in HCC-implications for screening. J Hepatol 2014;60:S45-S46.

44. Johnson PJ, Berhane S, Kagebayashi C, Satomura S, Teng M, Reeves HL, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. J Clin Oncol 2015;33:550-558.

45. Pinato DJ, Sharma R, Allara E, Yen C, Arizumi T, Kubota K, et al. The ALBI grade provides objective hepatic reserve estimation across each BCLC stage of hepatocellular carcinoma. J Hepatol 2017;66:338-346.

46. Toyoda H, Lai PB, O'Beirne J, Chong CC, Berhane S, Reeves H, et al. Long-term impact of liver function on curative therapy for hepatocellular carcinoma: application of the ALBI grade. Br J Cancer 2016;114:744-750.

47. Cho WR, Hung CH, Chen CH, Lin CC, Wang CC, Liu YW, et al. Ability of the post-operative ALBI grade to predict the outcomes of hepatocellular carcinoma after curative surgery. Sci Rep 2020;10:7290.

48. Hiraoka A, Kumada T, Tsuji K, Takaguchi K, Itobayashi E, Kariyama K, et al. Validation of modified ALBI grade for more detailed assessment of hepatic function in hepatocellular carcinoma patients: a multicenter analysis. Liver Cancer 2019;8:121-129.

49. Clavien PA, Lesurtel M, Bossuyt PM, Gores GJ, Langer B, Perrier A. Recommendations for liver transplantation for hepatocellular carcinoma: an international consensus conference report. Lancet Oncol 2012;13:e11-e22.

50. Lingiah VA, Niazi M, Olivo R, Paterno F, Guarrera JV, Pyrsopoulos NT. Liver transplantation beyond milan criteria. J Clin Transl Hepatol 2020;8:69-75.

51. Mehta N, Heimbach J, Harnois DM, Sapisochin G, Dodge JL, Lee D, et al. Validation of a risk estimation of tumor recurrence after transplant (RETREAT) score for hepatocellular carcinoma recurrence after liver transplant. JAMA Oncol 2017;3:493-500.

52. Hoffman D, Mehta N. Recurrence of hepatocellular carcinoma following liver transplantation. Expert Rev Gastroenterol Hepatol 2021;15:91-102.

53. Lee DD, Sapisochin G, Mehta N, Gorgen A, Musto KR, Hajda H, et al. Surveillance for HCC after liver transplantation: increased monitoring may yield aggressive treatment options and improved postrecurrence survival. Transplantation 2020;104:2105-2112.

54. 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;18:1206-1213.

55. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 2018;68:723-750.

56. Agopian VG, Harlander-Locke M, Zarrinpar A, Kaldas FM, Farmer DG, Yersiz H, 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;220:416-427.

57. Halazun KJ, Najjar M, Abdelmessih RM, Samstein B, Griesemer AD, Guarrera JV, et al. Recurrence after liver transplantation for

hepatocellular carcinoma: a new MORAL to the story. Ann Surg 2017;265:557-564.

58. Kim SH, Moon DB, Park GC, Lee SG, Hwang S, Ahn CS, et al. Pre-operative prediction score of hepatocellular carcinoma recurrence in living donor liver transplantation: validation of SNAPP score developed at Asan Medical Center. Am J Transplant 2021;21:604-613.

59. Creasy JM, Sadot E, Koerkamp BG, Chou JF, Gonen M, Kemeny NE, et al. Actual 10-year survival after hepatic resection of colorectal liver metastases: what factors preclude cure? Surgery 2018;163:1238-1244.

60. de Jong MC, Pulitano C, Ribero D, Strub J, Mentha G, Schulick RD, et al. Rates and patterns of recurrence following curative intent surgery for colorectal liver metastasis: an international multi-institutional analysis of 1669 patients. Ann Surg 2009;250:440-448.

61. Fong Y, Fortner J, Sun RL, Brennan MF, Blumgart LH. Clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer: analysis of 1001 consecutive cases. Ann Surg 1999;230:309-318; discussion 318-321.

62. He Y, Ong Y, Li X, Din FV, Brown E, Timofeeva M, et al. Performance of prediction models on survival outcomes of colorectal cancer with surgical resection: a systematic review and meta-analysis. Surg Oncol 2019;29:196-202.

63. Spelt L, Nilsson J, Andersson R, Andersson B. Artificial neural networks--a method for prediction of survival following liver resection for colorectal cancer metastases. Eur J Surg Oncol 2013;39:648-654.

64. Bilici A. Prognostic factors related with survival in patients with pancreatic adenocarcinoma. World J Gastroenterol 2014;20:10802-10812.

65. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, et al. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. JAMA Oncol 2017;3:1335-1342.

66. Falconi M, Eriksson B, Kaltsas G, Bartsch DK, Capdevila J, Caplin M, et al. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. Neuroendocrinology 2016;103:153-171.

67. Vagefi PA, Razo O, Deshpande V, McGrath DJ, Lauwers GY, Thayer SP, et al. Evolving patterns in the detection and outcomes of pancreatic neuroendocrine neoplasms: the Massachusetts General Hospital experience from 1977 to 2005. Arch Surg 2007;142:347-354.

68. Bar-Moshe Y, Mazeh H, Grozinsky-Glasberg S. Non-functioning pancreatic neuroendocrine tumors: surgery or observation? World J Gastrointest Endosc 2017;9:153-161.

69. Genç CG, Jilesen AP, Partelli S, Falconi M, Muffatti F, van Kemenade FJ, et al. A new scoring system to predict recurrent disease in grade 1 and 2 nonfunctional pancreatic neuroendocrine tumors. Ann Surg 2018;267:1148-1154.

70. Zou S, Jiang Y, Wang W, Zhan Q, Deng X, Shen B. Novel scoring system for recurrence risk classification of surgically resected G1/2 pancreatic neuroendocrine tumors - Retrospective cohort study. Int J Surg 2020;74:86-91.

71. He L, Li H, Cai J, Chen L, Yao J, Zhang Y, et al. Prognostic value of the Glasgow prognostic score or modified Glasgow prognostic score for patients with colorectal cancer receiving various treatments: a systematic review and meta-analysis. Cell Physiol Biochem 2018;51:1237-1249.

72. Zhang H, Ren D, Jin X, Wu H. The prognostic value of modified Glasgow Prognostic Score in pancreatic cancer: a meta-analysis. Cancer Cell Int 2020;20:462.

73. Liu Z, Jin K, Guo M, Long J, Liu L, Liu C, et al. Prognostic value of the CRP/Alb ratio, a novel inflammation-based score in pancreatic cancer. Ann Surg Oncol 2017;24:561-568.

74. Zhang K, Gao HF, Mo M, Wu CJ, Hua YQ, Chen Z, et al. A novel scoring system based on hemostatic parameters predicts the prognosis of patients with advanced pancreatic cancer. Pancreatology 2019;19:346-351.

75. Benavides M, Antón A, Gallego J, Gómez MA, Jiménez-Gordo A, La Casta A, et al. Biliary tract cancers: SEOM clinical guidelines. Clin Transl Oncol 2015;17:982-987.

76. Balachandran P, Agarwal S, Krishnani N, Pandey CM, Kumar A, Sikora SS, et al. Predictors of long-term survival in patients with gallbladder cancer. J Gastrointest Surg 2006;10:848-854.

77. Hawkins WG, DeMatteo RP, Jarnagin WR, Ben-Porat L, Blumgart LH, Fong Y. Jaundice predicts advanced disease and early mortality in patients with gallbladder cancer. Ann Surg Oncol 2004;11:310-315.

78. Bartlett DL, Fong Y, Fortner JG, Brennan MF, Blumgart LH. Long-term results after resection for gallbladder cancer. Implications for staging and management. Ann Surg 1996;224:639-646.

79. Fong Y, Wagman L, Gonen M, Crawford J, Reed W, Swanson R, et al. Evidence-based gallbladder cancer staging: changing cancer staging by analysis of data from the National Cancer Database. Ann Surg 2006;243:767-771; discussion 771-774.

80. Tran TB, Norton JA, Ethun CG, Pawlik TM, Buettner S, Schmidt C, et al. Gallbladder cancer presenting with jaundice: uniformly fatal or still potentially curable? J Gastrointest Surg 2017;21:1245-1253.

81. Cubertafond P, Gainant A, Cucchiaro G. Surgical treatment of 724 carcinomas of the gallbladder. Results of the French Surgical Association Survey. Ann Surg 1994;219:275-280.

82. Shukla PJ, Neve R, Barreto SG, Hawaldar R, Nadkarni MS, Mohandas KM, et al. A new scoring system for gallbladder cancer (aiding treatment algorithm): an analysis of 335 patients. Ann Surg Oncol 2008;15:3132-3137.

83. Leon AR. A new scoring system for gallbladder cancer: the first step of a long walk. Ann Surg Oncol 2008;15:2991-2992.

84. Wang K, Zhang H, Xia Y, Liu J, Shen F. Surgical options for intrahepatic cholangiocarcinoma. Hepatobiliary Surg Nutr 2017;6:79-90.

85. Pan QX, Su ZJ, Zhang JH, Wang CR, Ke SY. Glasgow Prognostic Score predicts prognosis of intrahepatic cholangiocarcinoma. Mol Clin Oncol 2017;6:566-574.

86. Wang Y, Li J, Xia Y, Gong R, Wang K, Yan Z, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. J Clin Oncol 2013;31:1188-1195.

87. Raoof M, Dumitra S, Ituarte PHG, Melstrom L, Warner SG, Fong Y, et al. Development and validation of a prognostic score for intrahepatic cholangiocarcinoma. JAMA Surg 2017;152:e170117.

88. Hahn F, Müller L, Mähringer-Kunz A, Schotten S, Düber C, Hinrichs JB, et al. Risk prediction in intrahepatic cholangiocarcinoma: direct

comparison of the MEGNA score and the 8th edition of the UICC/AJCC Cancer staging system. PLoS One 2020;15:e0228501.

89. Schnitzbauer AA, Eberhard J, Bartsch F, Brunner SM, Ceyhan GO, Walter D, et al. The MEGNA score and preoperative anemia are major prognostic factors after resection in the German intrahepatic cholangiocarcinoma cohort. Ann Surg Oncol 2020;27:1147-1155.

90. Aakre CA, Dziadzko MA, Herasevich V. Towards automated calculation of evidence-based clinical scores. World J Methodol 2017;7:16-24.

91. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ 2013;346:e5595.

92. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis research strategy (PROGRESS) 2: prognostic factor research. PLoS Med 2013;10:e1001380.

93. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10:e1001381.

94. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified

medicine research. BMJ 2013;346:e5793.

95. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. Circulation 2015;131:211-219.

96. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2014;13:8-17.

97. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12-22.

98. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019;393:1577-1579.

99. Dou D, Yang S, Lin Y, Zhang J. An eight-miRNA signature expression-based risk scoring system for prediction of survival in pancreatic adenocarcinoma. Cancer Biomark 2018;23:79-93.

100. Pencina MJ, D'Agostino RB Sr. Evaluating discrimination of risk prediction models: the C statistic. JAMA 2015;314:1063-1064.